DOCUMENT RESUME

ED 121 839 95 TM 005 271

AUTHOR Foldmesser, Robert A.
TITLE The Use of Test Scores as a Basis for Allocating Educational Resources: A Synthesis and Interpretation of Knowledge and Experience. Final Report.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO ETS-PR-75-25
PUB DATE Nov 75
GRANT NIE-G-74-0066
NOTE 116p.

EDRS PRICE MF-$0.83 HC-$6.01 Plus Postage
DESCRIPTORS *Compensatory Education; Educational Finance; *Educationally Disadvantaged; *Educational Resources; Elementary Secondary Education; Federal Aid; Low Achievers; *Resource Allocations; School Districts; Scores; Socioeconomic Status; State Aid; State Federal Support; Student Evaluation; Testing Problems; *Test Results
IDENTIFIERS *Test Based Funding

ABSTRACT

This study was undertaken to explore the issues involved in test-based-funding (TBF)--the concept that extra resources should be allocated to states or school districts on behalf of pupils with low test scores. The data for the study come from literature bearing on the concept, from documents in three states (Michigan, California, and New York) where it has been tried, and from interviews with key people who have been engaged in its development and implementation. In the first chapter, the origins and rationale of test-based funding is described and the problems with it that have been evident from the outset are briefly discussed. Chapter 2 describes the experiences of the three states with TBF. Of the three, Michigan has given the concept the greatest prominence in its state-aid program, and it occupies most of the attention, but the experiences of all three states are outlined. Since there are a number of questions connected with TBF on which the experience to date for one reason or another casts no light at all, Chapter 3 considers some of them, bringing to bear both logical analysis and such pertinent literature as there may be, and suggesting lines of future inquiry which may improve the understanding of the premises and implications of TBF. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED121839

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PR-75-25

THE USE OF TEST SCORES AS A BASIS FOR ALLOCATING EDUCATIONAL RESOURCES: A SYNTHESIS AND INTERPRETATION OF KNOWLEDGE AND EXPERIENCE

Robert A. Feldmesser

Final report of a study conducted
for the National Institute of Education
under Grant No. NIE-G-74-0066

Educational Testing Service
Princeton, New Jersey 08540
November, 1975

TM005 271

2

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

THE USE OF TEST SCORES AS A BASIS FOR ALLOCATING EDUCATIONAL RESOURCES: A SYNTHESIS AND INTERPRETATION OF KNOWLEDGE AND EXPERIENCE

Robert A. Feldmesser

Final report of a study conducted
for the National Institute of Education
under Grant No. NIE-G-74-0066

Educational Testing Service
Princeton, New Jersey 08540
November, 1975

The research reported herein was performed under a grant from the National Institute of Education. Grantees are encouraged to express freely their professional judgment in reporting on their research. Points of view or opinions stated in this report do not, therefore, necessarily represent the official position or policy of the National Institute of Education or of the Department of Health, Education, and Welfare.

TABLE OF CONTENTS

	Page
Acknowledgments	v
Chapter 1: Test-Based Funding -- Origins and Initial Apprehensions	1
Income-based funding	1
An alternative to IBF	6
Proposals and problems	16
Summary	22
Chapter 2: Experience with Test-Based Funding	25
Michigan	25
Basic provisions	26
Reactions to the test results	31
Efforts at evaluation	36
Operational problems and future prospects	42
California	47
New York	59
Summary and inferences	65
Chapter 3: Issues for Further Research	71
Entry testing	71
Criterion-referenced testing	75
The assumptions of TBF	80
The data base for IBF	92
Summary	97
References	100

TABLES

Table 2.1. Potential 1972-73 Allocations under Michigan's Section 3 Program	33
Table 2.2. Potential 1973-74 Allocations under Michigan's Chapter 3 Program	37

ACKNOWLEDGMENTS

I owe a heavy debt to the many staff members of the state departments of education in California, Michigan, and New York who gave generously of their time to answer my questions and to patiently explain the complexities of their state aid formulas. I hope they find that the information they gave me has been accurately presented and well used here.

The study benefited from the attention of an advisory committee, whose members were Fred G. Burke, Commissioner of Education for the State of New Jersey; Benjamin J. Henley, formerly deputy superintendent and acting superintendent of the public school system of the District of Columbia and a past president of the Washington Urban League; and Richard S. Hodes, chairman of the Education Committee of the Florida House of Representatives. They read a draft of the report and made a number of valuable suggestions for its improvement. An earlier version was read by several friends and colleagues, including Joel S. Berke, Nancy L. Bruno, Paul B. Campbell, Henry S. Dyer, John J. Fremer, Jr., James A. Kelly, and Iris Rotberg; their comments, too, were very helpful. Probably to my own disadvantage, I did not always take the advice of these discerning people, so the responsibility for the defects that remain in the report is, of course, mine alone.

Elsa Rosenthal was an industrious research assistant for the study, and Irma Kienitz an indefatigable secretary and typist. Their competence and energy were indispensable.

Throughout the conduct of the study and the preparation of the report, Richard A. Lallmang, the project officer at the National Institute of Education, showed great forbearance toward an author who seemed always to be behind schedule.

R. A. F.

THE USE OF TEST SCORES AS A BASIS FOR ALLOCATING EDUCATIONAL RESOURCES: A SYNTHESIS AND INTERPRETATION OF KNOWLEDGE AND EXPERIENCE

Report of a study made for the National Institute of Education
under Grant No. NIE-G-74-0066

Robert A. Feldmesser
Educational Testing Service

November 1975

Abstract

CHAPTER 1: TEST-BASED FUNDING -- ORIGINS AND INITIAL APPREHENSIONS

Pp. 1-6 Allocations of funds to support compensatory education have typically been based on a measure of socioeconomic status (SES), usually the income of families resident in the funded unit. Larger allocations are made to units with larger proportions of low-income families. This procedure, here called income-based funding (IBF), has been followed in several state programs and in Title I of the federal Elementary and Secondary Education Act.

Pp. 6-15 Two objections have been made to IBF: (1) The availability of data on family income lags behind changes in families' place of residence. (2) It is an inefficient way of allocating educational funds, in that many low-income children apparently do not need compensatory programs while some children who apparently do need them are not from low-income families. The premise of this latter objection is that a low level of school performance is itself the true indicator of educational need; low SES is only a proxy measure of it. Since a low test score is direct evidence of a low level of performance, it has been proposed that IBF be replaced by test-based funding (TBF). The idea is reinforced by other potential advantages: TBF would promote "accountability for results," would be more likely than IBF to equalize learning outcomes among children of different racial and socioeconomic groups, and would generate increased public support for larger appropriations of educational funds.

Pp. 16-24 At the federal level, H. R. 5163 was introduced as a TBF bill to replace Title I. Though it was not passed, extensive hearings on it were held and a residual appeared in the Education Amendments of 1974. TBF proposals were also made at the state level, and California, New York, and Michigan have had versions of it in effect for several years. Certain difficulties in TBF were recognized at the outset: (1) It would seem to have a "disincentive effect": the lower the scores of students were, the more money a district would receive, which might motivate educators to try to depress pupil performance instead of improving it. (2) It might

impel educators to manipulate scores so as to produce the appearance of low performance. (3) It might raise questions about the validity, reliability, and cultural bias of the test that was used, or provoke opposition if any standardized or norm-referenced test were used. Opinions differ as to the seriousness of these problems, but various methods of dealing with them have been tried or suggested.

CHAPTER 2: EXPERIENCE WITH TEST-BASED FUNDING

Pp. 25-31 Michigan enacted a TBF program, known as Chapter 3, in 1971. Under this program, for which $22,500,000 were appropriated in the first year, 67 districts received $200 in 1971-72 for each K-6 pupil estimated to be at or below the 15th percentile on three tests of the state assessment battery administered early in 1971. These pupils were assigned to reading or mathematics programs, which, however, could include other pupils as well. As a way of countering the disincentive effect, the districts were to receive the $200 in the two subsequent years only for those Chapter 3 pupils who, in the preceding year, had made at least 75 percent of a normal year's gain as measured on standardized pretests and posttests; for pupils making smaller gains, the amount was to be reduced proportionately, down to zero for pupils making no gain.

Pp. 31-36 Although pretests and posttests were administered in 1971-72, the gain-based funding provision was waived for the following year because appropriations had been enacted so late. The test results showed that, if the provision had been enforced, the districts would have lost $8,000,000 of their original allocations, and demands were consequently made that the provision not be enforced for the next year, either. The Michigan Department of Education acknowledged that reduction of funds would mean a loss of services to the pupils who were presumed to need them. A new section, 39a, was added to Chapter 3, enabling districts to re-earn virtually all the funds they would have lost, by promising to install a new "delivery system" for pupils who had made low gains. The 1972-73 test results showed a potential loss to the districts of nearly $5,000,000, and Section 39a was renewed for 1974-75.

Pp. 36-42 It is impossible to say whether Chapter 3 has achieved its aims. Data on test-score gains are not convincing. Moreover, if the gains were greater than would otherwise be expected, the improvement may have been attributable simply to increased funding, in which case a later reduction of funds would have been counterproductive; yet the improvement could not be attributed to an anticipated loss of funds, because it is not clear that local administrators and teachers could or did do anything they would not have done, anyway. Nor can it be said that Chapter 3 improved efficiency in the use of funds, because the target pupils were not segregated from others and because districts continued to receive their allocations for pupils whose scores did not improve -- as they would have to if the principle were to be upheld that test scores are a valid indicator of need.

Pp. 42-47 The implementation of Chapter 3 was attended by several operational problems, growing out of the effort to avoid regression in the measurement of gain, the concern in local districts that pupils not be given a pretest that was too difficult for them, and the fact that the tests were administered and scored by local school-district personnel. Also, local administrators complained that Chapter 3 was exclusively punitive, since it could lead to a reduction of funds but not to an increase. To meet this complaint, however, would imply abandonment of the TBF concept. The most likely outcome of legislative deliberations over the future of Chapter 3 is that TBF will be continued, with the same appropriation as at the start, but with a new allocation based on a more recent test administration; and instead of a requirement for pupil gain, districts will probably have to obtain departmental approval for their compensatory programs.

Pp. 47-55 California has had two statutes with TBF provisions. The first, the Miller-Unruh Basic Reading Act of 1965, was never funded at a level sufficient to allow the TBF provision to take effect. The second, the Early Childhood Education (ECE) act, became operational in 1973, with an appropriation of $25,000,000. ECE is not primarily a compensatory-education program; any district can qualify, and most of them have done so. Districts with plans approved by the State Department of Education receive a grant of $140 for each pupil in a participating school and grade, and an additional $70 for each pupil, up to 25 percent of average daily attendance, who is determined as having an educational need. The Department has defined pupils with need as being those who score in the lowest quartile of a statewide test in reading and mathematics. This definition was more a matter of expediency than of principle; test-score data were readily available and were believed to be simpler than other measures of need. In its first year, a district's plan was to be implemented in only some of its grades and/or schools. Subsequently, additional funds would be granted for expansion of the plan into other grades and schools upon determination by the Department that the district had so far met the objectives of its plan. Failing such a determination, funding would be continued without increase or decrease, except that a district would be considered for termination if it was refused expansion funds for two successive years. Beginning with the second year, decisions about expansion funds were to be based, to a limited extent, on attainment of pupil-performance objectives. These provisions are a functional equivalent to the requirement for pupil gain in Michigan, although test scores play a much smaller part in ECE than in Chapter 3.

Pp. 53-59 The burden on the districts for developing acceptable plans, and on the Department for reviewing these plans and determining whether objectives had been met, proved to be very great, and the procedures were later simplified. The Department is also considering proposals to eliminate the pupil-performance part of the expansion-fund decision, because of the potential disincentive effect. Only ten small districts were refused expansion funds two years in a row, and they will probably not be terminated. Data on test-score gains in 1973-74 were inconclusive, but it may be premature to judge the success of ECE on that basis. The TBF provisions of ECE have created less severe problems than did those in Michigan, chiefly because they have played so minor a role. Also, unlike Chapter 3, ECE offers an opportunity for a district to obtain additional funds beyond its original allocation.

Pp. 59-65 New York established, in 1967, an Urban Education program, under which supplementary funds were granted to 29 large districts with high proportions of pupils who had low scores on the 1967 PEP tests. In 1974, this program was replaced by PSEN (Pupils with Special Educational Needs), which gives a weight of 1.25 in the state's general aid formula to the average of the proportions of a district's pupils who scored below the fourth stanine on the PEP reading and mathematics tests in 1971 and 1972. This was regarded simply as a means of getting money to all districts that needed it, instead of to large districts only, and test scores rather than other measures of need were used because they were more uniform and convenient. Third-grade scores had been proposed, in the belief that their use would minimize the disincentive effect, but sixth-grade scores were adopted, out of the desire to refrain from influencing schools to adopt instructional methods that would produce low scores in the early grades. While departmental approval is required for PSEN projects, the approval process is not stringent, and up to 30 percent of the funds may be spent on otherwise ineligible pupils if necessary to avoid segregating target pupils. TBF has been less troublesome in New York than in Michigan or California, probably because few strings are attached to the allocated funds.

Pp. 65-70 This study of experience with TBF so far is not conclusive, but it suggests the following inferences: (1) The desire to avoid score manipulation seems to conflict with the objective of keeping the data base for funding up-to-date. (2) TBF may not improve efficiency in the use of funds, because the precise targeting of funds to the low-scoring pupils presumed to need them has been found to be administratively awkward or pedagogically undesirable. (3) The effectiveness of TBF in raising test scores has not been demonstrated, but it may be too early to expect that. Meanwhile, a dilemma has emerged in deciding what to do if test scores do not increase. Reducing funds would mean not meeting needs or denying that scores are need indicators; not reducing funds would prolong ineffective expenditures. (4) The ability of TBF to produce larger appropriations of funds for education is doubtful, at least at the state level. (5) Efforts to prevent a disincentive effect and the manipulation of test scores have not been entirely auspicious.

CHAPTER 3: ISSUES FOR FURTHER RESEARCH

Pp. 71-74 Analysis of some of the as yet untried solutions to problems of TBF, and of some of the assumptions associated with it, reveals areas that demand further thought and empirical investigation. Testing upon entry to school rather than in later grades may only change the way in which the disincentive effect is manifested instead of eliminating it. Manipulation of scores may be easier than it has seemed. The desire of teachers and administrators to demonstrate their competence by producing high scores might be sufficient to counteract the disincentive effect and the temptation to manipulate scores, but evidence is lacking.

Pp. 75-80 There is little reason to believe that criterion-referenced tests would overcome the opposition to TBF that it is thought may arise from the use of norm-referenced tests. Criterion-referenced tests are just as subject

to cultural bias and to use in invidious comparisons as are norm-referenced tests. In fact, the two are not different kinds of tests so much as different ways of interpreting test scores; and even the interpretations are not radically distinct, since a criterion is of necessity related to a distribution of scores. Moreover, criterion-referenced interpretations are especially inappropriate to decision-making situations in which the number of pupils to be placed in a category must be limited, which seems to be the case for TBF. Finally, the use of criterion-referencing for TBF purposes might well bring upon it all the criticisms that have hitherto been directed at norm-referencing.

Pp. 80-85 Among the assumptions of TBF that require further exploration are that compensatory-education funds should be spent exclusively on students who have extra needs; that TBF is more likely to equalize the scores of different SES groups than IBF is; and, most importantly, that test scores are true indicators of the need for additional educational resources. The definition of "educational need" that is consistent with TBF describes it as a "discrepancy" or "deficit" relative to a desired condition such as normal performance, leading to the implication that the absence of a low score signifies the absence of need. This implication would probably be rejected by most people if it were applied to children with handicaps like deafness or residence at a great distance from school. Additional resources may have reduced the effects of their handicap, enabling them to achieve higher scores than they might otherwise; but if so, it would seem that these resources, rather than being withdrawn if the children reach the average, should continue to be provided as long as the handicap exists, in order to sustain their higher performance. Thus, a more generally acceptable definition of "need" would be "a requirement for extra resources to overcome the effects of a condition adverse to learning."

Pp. 86-92 Low SES can properly be viewed as the handicap of "deficient capital embodiment." In that light, it would itself be the source of needs, not merely a proxy measure of them. The low scores of low-SES children would be interpreted as the symptom of a widespread need, rather than as the condition creating the need. Low-SES children with average scores would still need additional funds, to allow improvements in areas of their education other than those measured by the test scores or to enable those of high ability among them to reach high, rather than simply average, levels of achievement. These results are necessary elements of the concept of educational equality. Because it is aimed at all low-scoring pupils, TBF would divert funds away from districts which have low-SES children with non-low scores toward districts which have high-SES children with low scores. Present evidence does not yield a clear answer to the question of whether these funds could be expected to raise the scores of the children on whose behalf they would be allocated.

Pp. 92-99 The doubts about TBF lead to a reconsideration of IBF. The data base for it may not become obsolete as quickly as has been thought, and it could be kept up-to-date fairly easily. Parental education may be a more satisfactory measure of SES than family income, and a technique has been developed which would relate it to amounts of capital embodiment in children. It might contain its own kind of disincentive effect and an inducement for deception, but these would probably be less serious than with an income measure or with TBF. All of these propositions are also matters for further study.

CHAPTER 1: TEST-BASED FUNDING -- ORIGINS AND INITIAL APPREHENSIONS

This study was undertaken to explore the issues involved in "test-based funding": the concept that extra resources should be allocated to states or school districts on behalf of pupils with low test scores. The data for the study come from literature bearing on the concept, from documents in three states where it has been tried, and from interviews with key people who have been engaged in its development and implementation. In this first chapter, the origins and rationale of test-based funding will be described and the problems with it that have been evident from the outset will be briefly discussed.

Income-based funding

The costs of public elementary and secondary education in the United States have traditionally been paid largely out of funds raised by local school districts from sources within their own boundaries. As local revenue-raising capacities have become strained, however, state governments, and to a lesser extent the federal government, have stepped in to supplement local resources. Today, about 40 percent of the costs of public elementary and secondary education are paid out of state funds. In most states, the amount of the grant to a local school district is based, in one way or another, on the number of students enrolled in the district — an "equal dollars per scholar" formula -- sometimes adjusted to take account of differences in the districts' own revenue-raising abilities.

But for some time now, it has been recognized that there are important variations in the kinds and amounts of educational resources required by different sorts of children. A child who lives in a rural area, far from a school, needs more by way of transportation than does a child who lives nearer to one; a child with impaired vision or hearing needs more individual attention and specialized equipment and instructors, which call for greater expenditures than are required for ordinary children. A society which aspires to offer equal educational opportunity to all children, and to be just and humane, will try to meet these extra needs, and school districts in the United States generally do make that effort -- again, often with support from the state governments.

Children with special needs are not evenly or proportionally distributed among school districts, so it would not be satisfactory for the states simply to allocate additional funds to each district on the basis of its total enrollment. Either of two methods of determining the additional allocation is used instead. In the categorical-aid method, the state grants to the district a sum of money that can be used only to cover the extra expenditures required. Some states, for example, pay all or a substantial part of the costs of transporting children who live more than a specified distance from a school. Fewer states use the other method, pupil weighting, in which extra weight is given in the state aid formula to children who have certain recognized special needs; thus, a deaf or blind child may count for two ordinary children in computing the amount

of state aid a district will receive, without placing any restrictions on the district's use of the additional funds.

In recent years, it has been contended that children who are the victims of racial or ethnic discrimination -- particularly, in present circumstances, children socially defined as black, Hispanic, or Indian — and children from families of low socio-economic status (SES) also have special needs and therefore should have greater resources devoted to their education than other children. Minority and low-SES children are likely to suffer from educationally relevant disadvantages that are typically associated with their situation: the injuries to self-confidence, self-esteem, and academic motivation that result from prejudice and discriminatory treatment; inadequate nutrition, clothing, health care, and shelter and space; a restricted range of experiences; and peers and parents who provide little help in verbal development or other scholastic skills and who do not offer models of high aspirations.

Numerous studies have consistently shown that the average level of educational performance of children from minority and low-SES families is indeed significantly lower -- in both the technical statistical and the practically meaningful senses -- than that of children from families of the white majority and of higher SES. Many school districts have accepted the obligation to close these gaps. But, it is argued, the effort demands additional resources, to compensate for these disadvantages, just as in the case of the

other groups of children with special needs; and even more than these latter groups, children of minority and low-SES families are distributed among school districts in a highly skewed fashion, being concentrated in the large cities and in some rural areas. Once more, therefore, the districts have turned to higher levels of government for assistance. The federal and several state governments have responded; in some instances they have even acted first, offering additional funds as a stimulus to action by local districts that might not otherwise take the initiative.

For both judicial and political reasons, there is a widespread reluctance to allocate funds on the basis of membership in a racial or ethnic group. This reluctance seems not to apply, or at least not as strongly or as defensibly, to the allocation of funds on the basis of family income. Besides, additional funds directed toward children from low-income families can reasonably be expected to reach minority-group children in disproportionate numbers, because minority-group families tend to have lower incomes than majority-group families do. Consequently, the amount of additional funds provided has typically been determined by some measure of the income of families resident in the district, and the districts are expected or required to use the money for purposes of initiating and operating "compensatory" programs, designed specifically to raise the achievement level of low-income children. This procedure will be referred to here as income-based funding, or IBF.

Although federal aid constitutes only about seven percent of the total expenditures of public elementary and secondary schools, much of it is devoted to support for compensatory education. The largest single program of federal grants to local school districts is Title I of the Elementary and Secondary Education Act (ESEA), a program of categorical aid for compensatory education, for which amounts ranging between approximately one and two billion dollars have been appropriated in each year since 1965 (National Advisory Council on the Education of Disadvantaged Children, 1974, pp. 9, 71). Under the original provisions of Title I, a district was technically entitled to a grant equal to one-half the average per-pupil expenditure in its state (or in the nation, if that was greater) multiplied by the number of children of ages 5 to 17 in the district's population who were either from families with an annual income of less than $2,000 or from families that received more than $2,000 annually in welfare payments under the Aid to Families with Dependent Children (AFDC) program.[1] Since family-income and AFDC data were not usually available for school districts, the allocations were actually made at the county level, and the state

---

[1]The entitlement formula was later changed to reflect the number of families that were below the "poverty level" on an index developed by the Department of Commerce or that were receiving AFDC payments in excess of the poverty-level amounts. Also, the multiplier was changed from one-half to 40 percent of the state average per-pupil expenditure, except that it could not be less than 80 percent or more than 120 percent of the national average per-pupil expenditure.

educational agency (SEA) was then expected to distribute the funds to districts within counties on the basis of an estimate of the prevalence of low income or other socioeconomic characteristics in the districts. The districts received the money upon assurance to their SEA that they would use it "for programs and projects ... designed to meet the special educational needs of educationally deprived children..." In turn, the districts selected, as the ultimate recipients of the funds, those schools that were estimated to have the largest proportions of pupils of low SES. All pupils in these schools were eligible for Title I programs and projects, though not all of them necessarily participated. If the total Title I appropriation did not allow the full amount of the grant to be made to all districts -- and it never has been large enough for that -- the grant to each district was to be reduced proportionately.

An alternative to IBF

As experience accumulated with the state and federal programs of support for compensatory education, and as analyses were made of their implementation, dissatisfaction began to be expressed with the IBF concept. The criticisms were aimed particularly, though not exclusively, at Title I, because of the "leverage" it exerted on education for the disadvantaged (Berke and Kirst, 1972, p. 23), its national scope, and its special importance as a source of revenue for financially hard-pressed urban districts. Two major objections were made to Title I.

One of these pertained to the data used to determine the numbers of children on which the amounts of the allocations were based. Title I required, in effect, that the number of low-income children in each county was to be obtained from the U. S. Census, as the most reliable and uniform source of such information for all counties and as a way of ensuring that the size of a county's allocation was related to the magnitude of the needs of its districts. Consequently, the allocations were based on outdated figures -- the 1959 incomes ascertained in the 1960 Census. This was recognized even when the ESEA bill was being drafted, but the provision was nevertheless included in order to relieve school districts of the expense and awkwardness of having to gather their own income data by direct inquiry to resident families, since the Census tracts did not match district boundaries (Bailey and Mosher, 1968, p. 50). In later years, the obsolete nature of the data became ever more blatant. Congressman Albert H. Quie, one of the leading critics of the IBF provisions of Title I, described the situation in 1973 quite vividly:

> ... states are still being paid on the basis of a student who was enrolled 14 years ago. That means that if the student were in the second grade in 1960 he has probably graduated from high school, completed junior college and begun a family of his own. I see no defense for continuing to pay states as if that 21 year-old man or woman were still in the second grade (Quie, 1973, p. 3; emphasis in the original).[2]

---

[2]Actually, the 14-year interval was an artifact of the date of the passage of ESEA, midway between two Census years. In 1974, the basis of the entitlement formula was changed to the 1970 Census, and thereafter the numbers of low-income families would be at most "only" 10 years out-of-date.

AFDC data, supplied by the Welfare Administration from its payment records, were more nearly current, but had other kinds of difficulties. The definition of eligibility for AFDC varied from state to state and from time to time, and population groups differed in their propensity to apply for welfare assistance (Berke and Kirst, 1972, p. 108; Quie, 1974, pp. 3-4), leading to variations in the numbers of children qualifying for Title I funds that bore no necessary relationship to variations in the numbers of children needing compensatory education.[3]

A second criticism of Title I was that it was an inefficient way of allocating educational funds, in the sense that the money did not always find its way to the children who might have been thought to need it most. One indication of this came in an analysis by Gene V Glass of data collected by the Office of Education's Bureau of Elementary and Secondary Education on Title I programs in 1968-69. The teacher of each pupil in the Title I schools surveyed had been presented with a list of eight types of compensatory programs and had been asked, "According to your knowledge of this pupil's critical needs, which of the following would you recommend that he participate in during the next school year?" Glass found that, by this procedure, "Approximately 34 percent of the pupils were judged to have no

---

[3]State-level IBF programs of course face the same kinds of problems in using both income and AFDC data; see, for example, Benson and others (1972), p. 43; and Fleischmann Report (1973), vol. 1, p. 69.

critical needs for compensatory programs" (Glass and others, 1970, pp. 47-48). Teachers had also been asked to estimate the annual income of each pupil's family; more than 68 percent of the children in the lowest income category ($200-$499 per year per family member) had a critical need for reading, but 10 percent of them had no critical need in any of the eight areas, and although 64 percent of the pupils in the highest income category ($2600-$2800 per year per family member) had no critical needs, 19 percent of them did have a critical need in reading (Glass and others, 1970, p. 60). "Obviously," he concluded, "there are many children from families not suffering from poverty who have critical needs, and conversely, there are children from poor families who may not have needs in every area" (Glass and others, 1970, p. 61).

Congressman Quie calculated the figures in a different way and came up with a striking observation. Conceding that "there is unquestionably a high degree of correlation between poverty and educational deprivation," he pointed out that, nevertheless,

> ... we can extrapolate from the figures in the Glass study the fact that less than one-quarter of all students with educational deficiencies in the area of reading are in the below $2,000 family income classification (Quie, 1973, p. 2).[4]

[4]The assumption was made that the average family size was four, so that $499 per year per family member translated into an annual family income of $1996. The estimated 350,548 pupils who were in that income category and were judged to have a critical need in reading constituted 68.5 percent of all pupils in the category, but only 14.2 percent of all pupils with a critical need in reading. Even if the estimated 218,701 pupils with a critical need in reading for whom there was no income information were added to the number, the total would still make only 23.1 percent of all pupils with a critical need in reading. It should be borne in mind, however, that the survey covered pupils in Title I schools, not all of whom were participants in Title I programs or projects; and that Glass repeatedly emphasized in his report that the pupils in the survey were not necessarily a representative sample of pupils in Title I schools.

Thus, the income criterion was preventing Title I funds from reaching three-quarters of the children who needed help in reading. "The other side of the coin," added the congressman, "is that ... many low income students are simply not in need of compensatory education" (Quie, 1973, p. 2). He was charging, in other words, that Title I used a "buckshot" approach: It sprayed funds at school districts with large proportions of low-income children in the hope that the money would "hit" those children with the greatest educational needs. But there were many "misses" -- the money apparently reached some children who didn't need it and failed to reach some who did. Economist Charles Benson has used a different metaphor to make the same point about IBF at the state level:

> ... as long as classrooms are integrated racially, socially, and economically, it is inevitable that resource "leakage" occurs. That is, too much money leaks to students who are doing well enough by ordinary standards .... It follows that income-based grants are not finely adjusted instruments for attacking the problem of school failure ... (Benson and others, 1974, p. 85).[5]

Implicit in such positions is the premise that a low level of school performance is itself the true indicator of educational need. This premise had earlier been made explicit in a study by Garms and

---

[5] In an analysis of school financing that he had made earlier for the California State Senate, Benson argued that use of an AFDC measure was also inefficient, for a different reason: It "prevents us from weighting students according to degree of need because welfare status is a dichotomous variable" (Benson, 1972, p. 43). It may be noted that the income measure used in Title I was dichotomous, too, but that could have been easily remedied by graduating the size of the grant in accordance with the proportions of children from families of various incomes. Conceivably, the same could be done with amount of welfare assistance received.

Smith for the New York State Educational Conference Board, dealing with the allocation of educational funds in that state. The authors declared, in their introduction: "Educational need exists wherever average achievement levels are consistently and significantly below the norm" (Garms and Smith, 1969, p. 7; emphasis in the original). Moreover, they went on,

> The most direct measure of educational need ... would clearly be pupil achievement as indicated on [sic] test scores. Since the purpose of this study is to develop a way of distributing state aid which more accurately reflects educational need, the most obvious method would be to allocate funds in accordance with test results (Garms and Smith, 1969, p. 8).

The authors believed, however, that there were several unresolved problems that prevented use of that method for the time being -- problems which we will mention below -- and so they proceeded to search instead for "some measure or measures which correlate highly with student achievement" (Garms and Smith, 1969, p. 8) and which would be more easily available and more current than family income. They ended up with a measure that combined several other indices of socioeconomic status, but it was, as they said, a "second alternative."

The methodology that had been used by Garms and Smith in New York was subsequently applied in a study of the relationship between socioeconomic status and achievement in four other states, and the premise was restated in the report of the study, with an interesting corollary:

> We ... argue that educational need cannot be defined without reference to educational achievement, and that a need for compensatory education ... exists wherever there are consistent and significant differences in

> average levels of achievement among racial, economic, and social groups ... there would be no need for compensatory education if the correlation between socio-economic status and achievement were reduced to zero, or at least to an insignificant level (Garms and Kelly, 1970, pp. B10-B11).[6]

James A. Kelly, one of the authors of this study and a consultant to the Garms and Smith study, discussed the assumptions and findings of the latter, and its implications for school financing in New York City, in an article in Education and Urban Society, in which he wrote that socioeconomic measures were "a proxy for educational achievement" (Kelly, 1970, p. 266), and that term nicely encapsulates this criticism of IBF. Title I and similar state provisions were inefficient because they used a measure of educational need that was only indirect, a "proxy."[7] The evidence

---

[6]This study is summarized in Burke, Kelly, and Garms (1971), and its predecessor is summarized in Garms and Smith (1970).

[7]The term was used in the same way in two later reports. One, apparently written for internal circulation in the Office of Education, asserted (without supporting documentation) that "Congress ... has made it clear that the target groups [of Title I are] the educational under-achievers rather than low income children per se" and then went on:

> Unfortunately it is not possible to use a direct measure of the educationally disadvantaged because there are no uniform statistics on educational achievement currently available. Because of the unavailability of uniform performance measures and the close correlation between education and income, low income is used [in Title I] as a proxy for the educationally deprived (Wilensky, 1972, p. 3).

The other was the report of a study of school finance in eight states done at The Urban Institute:

> A major concern is to identify precisely the particular pupils who need special help. Available data do not provide consistent or generally acceptable measures for identifying such pupils, so that reliance is frequently placed on membership in particular socio-economic or racial groups as proxies (Levin and others, 1972, p. 203).

that extra resources were needed for the education of low-income children was that their test scores were lower than those of other children; but if that was the case, it seemed logical to do away with the proxy and to try to channel the resources directly to low-scoring children to begin with, instead of using the circuitous route of family income. IBF should therefore be replaced, in whole or in part, by TBF--test-based funding.

Converging with this line of thinking was an increasing emphasis in many quarters on "accountability for results" in education. The idea had been gaining currency that the quality of an educational system was not to be judged by its teacher-pupil ratio or the age of its buildings or other traditional "input" criteria, but rather by the system's "output" -- the performance of its students. The clearest evidence of student performance was test scores, and if fund allocations were to depend on test scores, the scores would obviously have to be made public, thus calling attention to this fundamental dimension of educational operations. The results would be, as James S. Coleman (1971) among others argued, that pressure would be brought to bear on teachers and administrators to concern themselves with student learning rather than with less important matters. Indeed, the mere "public feedback" of test results to teachers might serve as an effective incentive for them to improve the results of their work (Lipe and Jung, 1971).

The growing stress on performance outcomes had another, especially pertinent aspect. In legislative debate, judicial litigation, and scholarly writings on school finance, the argument was being made that the success of an educational system in providing equal treatment to all who passed through it was also to be measured by outcomes among students: by whether children of different racial and socioeconomic groups were, on the average, learning to read and calculate equally well. If they were not, the lower-scoring children were being deprived of their rights and the system was obliged to remedy the situation. This led easily to the further argument that the most likely way of eliminating educational inequalities among racial and socioeconomic groups was to provide additional resources to low-scoring children. As Garms and Smith (1969, p. 7) said:

> If educational need is defined in terms of education achievement and if resources are applied according to some measure of this need, differences in average achievement levels among different social, economic, and racial groups will hopefully decrease ...

In other words, targeting resources directly to low-scoring children would be not only a more efficient mode of allocation; it would also be more effective, in the sense that it offered greater promise of raising those low scores.[8]

Congressman Quie pointed to still another potential advantage of TBF. The reason why Title I had been consistently under-funded, he said (Quie, 1973, p. 10), is that "the only public support for

---

[8]For a similar view, see Quie (1973), p. 5.

it comes from the poorest elements of our society," since they are its major beneficiaries. But under a TBF formula, funds that had formerly gone to low-income children who were not low performers would be diverted to children of higher-income families who were low performers. The result would be that support for increased appropriations would be generated among those higher-income -- and politically more powerful -- families.

> If we begin counting [educationally] needy children, not poor children, we will create a program support base which will include virtually every parent and the relatives of every child who is in need of compensatory services. The pressure from those people to move the funding levels upwards will be enormous (Quie, 1973, p. 10).[9]

At the same time, low-income families would not object, because even though they received a smaller share of the pie, aid for their children's education would not decrease (or might even increase), since the pie would be larger.

---

[9] In more informal remarks made a year later, the congressman put the point in blunter and more colorful terms:

> It is interesting that so many educators talk about increasing federal funding for elementary and secondary schools and yet we continue to count only the children of parents without any political clout. In the past, certainly families with $2,000 in income and AFDC had very little political clout. We leave out the other large number of educationally disadvantaged children [i.e., those not performing well] who come from families that do have political clout. So, just from the politics of it, you would think we should count everyone who needs compensatory education so that people with some political force behind them could say, "Hey, we want more money for our kids!" (Quie, 1974, p. 6).

Proposals and problems

The TBF bill which Congressman Quie introduced in 1973, H. R. 5163, was intended as a replacement for Title I of ESEA. It called for the establishment of a National Commission on Educational Disadvantage, which was to develop a test "to measure the performance of children in terms of specific criteria ... of what children should know or be able to do at selected age or grade levels" in reading and mathematics. The test would be administered biennually, the Commission would set a "standard of performance" on it, and children who failed to meet the standard would be defined as "educationally disadvantaged." Each state would then receive a grant of funds equal to 40 percent of the per-pupil expenditure in the country as a whole or in that state in particular, multiplied by the number of children 5 to 17 years of age who were educationally disadvantaged by that definition, or who were students in state-supported schools for the handicapped or for neglected or delinquent children or who were the children of migratory agricultural workers. The states were to distribute these funds to local school districts for programs "designed to meet the special educational needs of educationally disadvantaged children ..." The bill did not get out of committee, but extensive hearings on it were held and a residual of it did appear in the Education Amendments of 1974, which, in extending the ESEA to 1978, authorized experiments in up to 20 districts which would, among other things, explore "alternative methods, including the use of procedures to assess educational disadvantage, for distributing funds ..."

Meanwhile, there had been several moves toward TBF at the state level. In 1972, Benson proposed to the California Senate that, while half of the state's funds for compensatory education continue to be allocated on the basis of SES, the other half depend on student test scores. He summarized his proposal thus:

> The amount of aid allocated to a district on behalf of any student would be determined by (i) the difference between the student's SES and the mean for the state and (ii) the difference between his grade one test score and the mean for the state. Furthermore, the larger these differences are, the higher is the amount of aid which is allocated to the student. Funds to the district as a whole would simply be the sum of the allocation[s] for individual students in the district (Benson, 1972, p. 40).[10]

This proposal was not adopted,[11] but California did enact other, more modest versions of TBF in 1965 and 1971. Experience with these will be considered in some detail in the next chapter.

Also in 1972, the New York State Commission on the Quality, Cost, and Financing of Education (the Fleischmann Commission), following a wide-ranging investigation of elementary and secondary education in that state, recommended that, in order "to begin to regulate the flow of funds according to the educational needs of the students,"

---

[10]Though it is clear from the context that Benson intended aid to be higher as the student's test score was further <u>below</u> the mean for the state, he does not say whether the allocation to a district would take account of scores that were <u>above</u> the state mean.

[11]Nor was one which was reported to have been made to the Missouri legislature in about 1968 by Alan Thomas, who suggested (according to Kelly, 1970) that the state aid formula provide additional amounts to districts according to the proportion of pupils scoring below the 30th percentile on the distribution of scores on a statewide test.

> students who score at a low level in reading and mathematics achievement be weighted at 1.5 [in the state aid formula], as against a weighting of 1.0 for other children, and that the proportion of students so selected be based upon the proportion of third-grade students in each district who obtain marks at or below the third stanine on third-grade reading and mathematics achievement tests currently being administered in the state Pupil Evaluation Program ... (Fleischmann Report, 1973, vol. 1, p. 66; the entire passage was underscored in the original).[12]

A modified form of this recommendation was implemented in New York in 1974, replacing an earlier and much more limited TBF provision that had been passed in 1967. These, too, will be examined in the next chapter, along with Michigan's TBF program, enacted in 1971 and the most comprehensive embodiment of the concept so far.

Even while these proposals were being made and enacted, it was recognized that there were certain difficulties inherent in the logic of TBF. Probably the most serious of these was that it would seem to have a "disincentive effect": the lower the scores of students were, the more money a district would receive, which could be interpreted as a "reward" for doing an inferior educational job and might actually motivate educators to try to depress pupil performance instead of improving it. Conversely, aid funds would decline where test scores were rising, so that districts would lose funds "just at the time when they ... have discovered how to use them effectively" (Fleischmann Report, vol. 1, p. 68; see also Garms

---

[12]The California and New York proposals were not entirely independent; Charles Benson was study director for the Fleischmann Commission. However, the Regents of New York State also made a similar but less specific recommendation in 1972 (Reischauer and Hartman, 1973, p. 92).

and Smith, 1969, p. 7; Kelly, 1970, p. 266; Madaus and Elmore, 1973, pp. 3018-3019). If these consequences were to occur, they would defeat two of the hopes for TBF: that it would help to hold educators accountable for "results" and that it would increase effectiveness in the use of educational funds.

Three types of resolutions to this dilemma have been devised or suggested. The first is to base the amount of the grant initially on the number of low-scoring children but subsequently on the degree to which their scores increase. The second is to place tight controls on the uses to which the additional funds could be put, or even to provide the extra resources in the form of "outside help" for specified purposes (Garms and Smith, 1969, p. 80), so that at least they would not be viewed as a reward. The third is to administer the test which defines low performance to pupils so early in their educational careers that the schools would have little opportunity to depress the scores on them. The first of these was the approach adopted in Michigan, and the second in California; their success in coping with the problem will be discussed in the next chapter. The third has not been tried; its ramifications will be considered in the last chapter.

Even if TBF did not motivate educators to depress pupil performance, it might impel them to produce the appearance of low performance, and that would not seem to be good policy, either, since it would channel funds to those districts which were best at

dissembling. Congressman Quie acknowledged this possibility but did not think it very likely:

> It has been suggested that the education community would cheat if it got more money when its students were under-achievers. This would be a little difficult because the persons who seem to be in charge of getting the money are either the school superintendent or the school board, and they would then have to get the principals of all the schools to cooperate, and the principals, the teachers. Then the teachers would have to persuade all the students to do poorly to have the school district get more money (Quie, 1974, pp. 6-7).

The Fleischmann Commission was not as ready to dismiss the problem; it suggested that, in order to prevent the manipulation of scores, the test could be administered and scored by the state or "perhaps by an outside agent" (Fleischmann Report, vol. 1, p. 68). Another method, also recommended by the Commission (Fleischmann Report, vol. 1, p. 68) and utilized in both New York and Michigan, would be to base the grants on scores on a test administered before the TBF provisions were enacted.

Garms and Smith (1969, p. 8) alluded to one other potential difficulty in implementing TBF: "The use of a standard test for distribution of state aid would raise questions regarding the validity, reliability, and cultural bias of the tests employed." In some respects, these would not appear to be troublesome issues. A lack of validity and reliability would have random effects and would therefore mis-classify as many high-performing students in a low-scoring category as the reverse. As far as cultural bias

is concerned, the districts that would suffer by a shift from IBF to TBF, and that would consequently be most strongly inclined to "raise questions" about the test, would be districts with large proportions of pupils from low-income and minority-group families -- and ironically, it is just such districts that would be "favored" by a culturally biased test, since it would show their pupils scoring lower than the pupils' actual achievements warranted.

On the other hand, standardized, or "norm-referenced," tests have been subjected to much criticism in recent years for a variety of other reasons, many of them centering around the fact that these tests lend themselves to invidious comparisons among children and are believed thereby to interfere with the learning process. Hence, quite aside from technical problems of validity, reliability, or bias, it might be expected that TBF would arouse opposition if it were to be based on the scores on norm-referenced tests, since the distribution of large sums of money on that basis would give the tests greater importance than ever. A 1973 resolution of the National Education Association (1973, p. 87) urged "the elimination of group standardized intelligence, aptitude, and achievement tests ... until completion of a critical appraisal" of such tests, and indeed an interim report from the task force established to conduct this appraisal said that "[t]he results from group standardized tests should not be used as a basis for allocation of federal or state funds" (Today's Education, 1974).

Presumably to avert such opposition, Congressman Quie proposed, in H. R. 5163, that allocations be based instead on the scores on a "criterion-referenced" test -- i.e., one that, as we have seen above, would "measure the performance of children in terms of specific criteria ... of what children should know or be able to do at selected age or grade levels," thus comparing each child's performance to some desired performance rather than to the performance of other children. Although criterion-referenced testing is not intrinsic to the principle of TBF, much of the testimony on H. R. 5163 was concerned with its feasibility and advantages, and it also arose as an issue in the Michigan TBF program. We shall have more to say about it later.

Summary

Financial assistance by the federal and state governments to help local school districts pay for the extra costs of compensatory education has rested on the assumption that these extra costs arise out of the disadvantaged status of low-income and minority-group families. It has therefore generally taken the form of income-based funding (IBF), in which allocations are determined by the income of families resident in the district (or in its county). This procedure has been criticized particularly on two grounds: (1) The data on which it is based, family incomes as shown in Census publications, eventually become obsolete. (2) It is an inefficient way of allocating funds, because the true indicator of educational

need is the actual performance of pupils, and some children of low-income families are performing adequately while some children of high-income families are not.

Since performance can be measured by test scores, proposals have been made to replace IBF by TBF -- test-based funding, in which allocations would be determined directly by pupils' scores on tests of achievement in reading and mathematics, rather than by the "proxy" of their families' incomes. It is argued that, in addition to keeping the data base current and distributing extra funds to where they are most needed, TBF would promote educators' accountability for learning outcomes, would increase effectiveness in the use of funds and thus improve the chances for reducing score disparities among socioeconomic groups, and would strengthen support for an increase in total educational funding.

At the federal level, the Quie bill, H. R. 5163, was introduced in 1973 as a TBF replacement for the IBF provisions of Title I of ESEA. Extensive hearings on it were held, but it was not passed; however, the idea was kept alive in the Education Amendments of 1974, which authorized experimentation with it. TBF proposals were also made at the state level, and various forms of it have been enacted in Michigan, California, and New York.

Certain problems with TBF were apparent from the outset: (1) The potential "disincentive effect" of having a district's funding increase as the performance of its students declined, and decrease

as performance improved. (2) The possible encouragement of manipulation of scores to give the appearance of low performance. (3) The opposition that might be aroused by reliance on standardized or "norm-referenced" tests. Methods of dealing with these problems have been tried or suggested, and they will be discussed in the next two chapters.

CHAPTER 2: EXPERIENCE WITH TEST-BASED FUNDING

Three states -- Michigan, California, and New York -- have had versions of TBF in effect for varying numbers of years. Of the three, Michigan has given the concept by far the greatest prominence in its state-aid program, and so it will occupy most of our attention, but the experience of all three states is instructive.

Michigan[1]

Section 3 of Michigan's State School Aid Act has long provided support for compensatory education programs in the local school districts. For some years, the grants had been awarded competitively, the maximum amounts being based on various socioeconomic characteristics of the districts, such as the percent of children from families receiving welfare assistance or from broken homes, the percent of housing units undergoing clearance or rehabilitation, etc. In 1971, the section was revised so that the amount of state aid to a district for compensatory education would be determined by the test performance

---

[1]This account is based on a lengthy group interview and several individual interviews of half a dozen staff members of the Michigan State Department of Education, and on the following publications of the Department: A Description and Evaluation of Section 3 Programs in Michigan 1971-72: Report Number 1; A Description and Evaluation of Section 3 Programs in Michigan 1971-72: Report Number 2; A Description and Evaluation of Chapter 3 State Compensatory Education Programs in Michigan 1972-73: Part One; and A Description and Evaluation of Chapter 3 State Compensatory Education Programs in Michigan 1972-73 [Part Two]. These publications will be cited by year and report or part number. There is an excellent discussion of the Michigan TBF program, and of its relationship to the state's accountability efforts, in Murphy and Cohen (1974).

of its students. This change was sought by the State Superintendent of Public Instruction, Dr. John W. Porter, and by the Michigan Department of Education (MDE) generally, as part of a broader thrust toward accountability. The arguments used to the legislature, and their results, were some of those which we have already seen offered for TBF, as attested to by these statements of departmental staff members:

> . . . this was a powerful argument that state educators were able to mount that had great appeal to state legislators: that here was direct evidence of need as measured by a test -- and they're willing to accept test scores. Furthermore, districts were going to gain or lose money based on their performance.
>
> . . . the idea of some kind of a program focused on schools with the greatest educational need that had a feature of accountability built into it freed up 13 million new dollars [the increase in Section 3 appropriations from 1970-71 to 1971-72] . . . . There was [among the legislators] a sentiment against sociological data — but they look at test data and see that as being hard and firm data, not the usual kind of stuff that comes out of ivory towers. They were quite willing to buy that as a real indication of need in Michigan schools.

Nevertheless, perhaps because of the legislators' uncertainty about its wisdom or efficacy, the new program was given an initial life of only three years.

Basic provisions

In the first year, the districts' allocations were to be based on estimates of the proportions of their students who were "low achievers." These estimates were to be derived from the results of the January, 1971, administration of the Michigan State Assessment Battery, a set of tests that had been specially developed as another part of the state's effort to institute district accountability and that were administered to pupils in grades 4 and 7. (Although the battery was given each year, use of the

January, 1971, results avoided the possibility of manipulation of scores.) The district's enrollment in grades K-4 was to be multiplied by its proportion of fourth-grade pupils whose composite score in reading, mechanics of written English, and mathematics was at or below the 15th percentile of the score distribution for the state as a whole; its enrollment in grades 5 and 6 was to be multiplied by the proportion of its seventh-grade pupils whose score was at or below the 15th percentile; and these two products were to be added together and expressed as a proportion of its K-6 enrollment. A district was eligible for Section 3 funds if the sum of the products -- i.e., the number of K-6 pupils thus estimated to be at or below the 15th percentile -- was at least 30 and constituted at least 15 percent of its K-6 enrollment. The district with the largest proportion was to receive $200 times the estimated number of low-achieving pupils; similarly for the district with the next largest proportion; and so on, until the total amount appropriated for the purpose, $22,500,000, had been completely allocated.

By this procedure, 67 districts (about half the number that were eligible) were designated as recipients of Section 3 funds for 1971-72; the allocations ranged from $11,853,400 for Detroit to $7000 for the town of Rock. The program was to be confined to these 67 districts for all three years of the statute's life, in order to allow for development and continuity of local programs. No restrictions were placed on the use of the funds except that they

had to be devoted to compensatory education programs in the basic cognitive skills for K-6 pupils other than those being funded under programs for the handicapped, and they could not be used for cross-district busing to achieve racial balance.

In 1971, when Section 3 was revised in this way, "performance contracts" were enjoying a wave of popularity. Under these contracts, an organization, typically a private firm, provided instructional services to selected students in a school district and was paid a fee which depended upon the amount of learning demonstrated by the students served. Because a district's Section 3 funding for the second and third years of the program was to be similarly dependent, MDE publications often refer to the act as a "performance contract" between the state and the local school districts.[2]

Each district receiving Section 3 funds was to identify the specific pupils in grades K-6 who would be the beneficiaries of these funds. The number of pupils was to be equal to the number by which the district had qualified for Section 3 funds, but of course the qualifying procedure could not be used to identify all the specific pupils. The act stated that "the pupils to be provided special assistance with these moneys" were to be "selected in grades 2-6 from the lowest achievers in basic cognitive skills and in grades K and 1 from among those with the lowest readiness for the acquisition of cognitive skills." In its implementing rules, the MDE defined

[2]In fact, a subsection of the act appropriated $500,000 for the more literal sort of performance contracts. For a discussion of Michigan's experience with these contracts, and of the history of performance contracting in general, see Feldmesser and Echternacht (1975).

"lowest achievers" in grades 2-6 to be those pupils who had actually scored at or below the 15th percentile on the state assessment battery in 1971 (applicable only to pupils who had then been in the fourth grade), or those pupils who had scored "1 or more years below grade level on a standardized achievement test" which they had taken on or before October 1, 1971. In the absence of either of these kinds of information, "the attested judgment of a school teacher or school official ... that the child is in need of substantial improvement in the basic skills" could be substituted. K-1 pupils with "lowest readiness" were defined as those for whom the "results of a standardized readiness instrument indicate the need for substantial improvement in readiness skills"; or, again, the judgment of a teacher or school official could be used if no test results were available, and that was the method used for selection of K and 1 pupils in the great majority of districts.[3] Pupils identified as Section 3 pupils remained such for three years or until they transferred out of the district or completed the sixth grade. Pupils "lost" to a district in either of these two ways could be replaced by others identified in the original manner. Section 3 pupils remained in the buildings and classrooms where they already were, but were assigned to a reading program, a mathematics program, or both. These programs could, and usually did, include non-Section 3 pupils as well.

---

[3]The use of teachers' judgments may have been handy when the number of pupils eligible by achievement-test criteria was smaller than the number the district was entitled to by the qualifying procedure. If the number of eligible pupils was larger than the entitlement number, districts were to select those furthest below grade level or judged to have the greatest need for improvement.

The performance of the Section 3 pupils in these programs was to determine the Section 3 funding of the districts for the second and third years; this was to be the technique for avoiding the disincentive problem of TBF. The act required that recipient districts specify "the performance objectives of their compensatory education program." Each year, "an assessment or evaluation of the progress" of the participating pupils was to be made "with the use of pretests and posttests." For each pupil "making a minimum gain during the year of at least 75% of the skills in the performance objectives specified for his program," the district would again receive $200 the following year. For pupils achieving less than 75 percent gain, the funding would be reduced proportionately; e.g., for a pupil who made a gain of 30 percent (= .4 of 75 percent), the district would receive $80 (= .4 of $200). A district would lose its $200 allocation entirely for each pupil with a zero gain or less. However, the full $200 would be retained for pupils who had not had 150 days of instruction because of illness or who had moved out of the district before the end of the school year.

The wording of these provisions -- "performance objectives" and "75% of the skills" -- sounds as if the legislature hoped that districts would measure pupil performance with criterion-referenced (or "objectives-referenced") tests rather than with norm-referenced tests. However, with a little semantic sleight of hand, the MDE, probably in recognition of the fact that criterion-referenced tests

were not readily available, opted for norm-referenced tests. Its implementing rules required a district to have but a single performance objective, "an increase in achievement equivalent to 1 year's growth," and "1 year's growth" meant that achievement had to be measured in grade-equivalent terms and therefore with norm-referenced tests. Instruments were to be "selected from a list of standardized [i.e., norm-referenced] achievement tests covering communication and computational skills approved by the department ..." Thus, in effect, a district would receive second- and third-year funding in proportion to the degree to which its Section 3 pupils gained three-quarters of a year in grade-equivalent units of either a composite score in reading and mathematics or a score in one or the other, depending on which program or programs they had participated in.

Reactions to the test results

Pretests and posttests for carrying out the provisions of Section 3 were duly administered at the beginning and end of the 1971-72 school year. However, funds for the Section 3 programs were not appropriated until October and did not begin flowing to the school districts until midway through the school year. Consequently, the MDE felt that it would be unfair to the districts to enforce the performance-based funding provisions in 1972-73, and at its request, the legislature postponed them for a year. In effect, the first year was "written off," and a new three-year cycle was to begin with the

1972-73 year. The recipient districts, their initial allocations, the pupils identified as participants (or their replacements), and all other provisions remained the same. The only change, other than the dates, was that Secion 3 became Chapter 3, the name by which Michigan's TBF program has been known since.

Following a preliminary report late in 1972,[4] the MDE published, in March, 1973, a report (1971-72: Report Number 2) on the gains of participating pupils during the program's first year. The figures indicated that, if the performance-based funding provisions had been enforced, the 67 districts would have lost some $8,000,000 of their original Section 3 allocations; for Detroit alone, the loss would have been about $5,000,000 (see Table 2.1). There were many weaknesses in the data; for example, adequate score reports were not available for 16 percent of Section 3 pupils, accounting for nearly half of the total "losses," and some of this may have come about merely because districts were not overly diligent about submitting test information once they knew that it was not going to be used for funding purposes (1971-72: Report Number 1, p. 19).[5] Nevertheless, it must have caused some forebodings to discover that, among pupils for whom scores had been reported, 16 percent had made no gains, and another 18 percent had made gains below the 75-percent level,

---

[4]This report (1971-72: Report Number 1) is chiefly important for its references to problems encountered in implementing the act, which will be discussed later.

[5]This probably explains why Detroit did not report any students moving out of the district or failing to receive 150 days of instruction because of illness. Most of these pupils were presumably included in the "data not available" category.

Table 2.1. Potential 1972-73 Allocations under Michigan's Section 3 Program

	Detroit		Other 66 districts		Total for 67 districts	
	Number of pupils	Potential allocation	Number of pupils	Potential allocation	Number of pupils	Potential allocation
Total for 1971-72	59,267	$11,853,400	53,233	$10,646,600	112,500	$22,500,000
Pupil category for 1972-73 funding						
Made gain of 75% or more	30,271	$ 6,054,200	31,865	$ 6,373,000	62,136	$12,427,200
Made gain of 50-74%[a]	3,605	$ 596,027	4,210	$ 696,053	7,815	$ 1,292,080
Made gain of 25-49%[a]	2,362	$ 233,050	2,914	$ 287,515	5,276	$ 520,565
Made gain of 1-24%[a]	1,765	$ 58,833	2,418	$ 80,600	4,183	$ 139,433
"Special eligibility"[b]	---	---	504	$ 100,800	504	$ 100,800
Subtotal	38,003	$ 6,942,110	41,911	$ 7,537,968	79,914	$14,480,078
% of 1971-72	64	59	79	71	71	64
Made no gain	7,452	($1,490,400)[c]	7,221	($1,444,200)[c]	14,673	($ 2,934,600)[c]
Data not available	13,812	($2,762,400)[c]	4,101	($ 921,000)[c]	18,417	($ 3,683,400)[c]

[a]Potential allocations for these categories were estimated by using the midpoint of the interval to represent gain for all pupils in the interval.

[b]Pupils who transferred out of the district prior to the posttest or who did not receive 150 days of instruction because of illness.

[c]These are amounts ($200 x number of students) which districts would not have received, in addition to losses due to pupils who made less than 75% gain.

Source: 1971-72: Report Number 2, pp. 13, 27, for numbers of pupils. Potential allocations were calculated for this study.

representing a potential loss of more than $4,000,000 in state aid. There may not even have been great confidence among district officials in their ability to reduce the number of pupils with inadequate data, for, as the Department's first report commented, "The mere task of reporting each individual student's record" -- and Detroit had almost 60,000 to report on -- "is a formidable task which required [a] considerable amount of discipline" (1971-72: Report Number 1, p. 18). Many of the pupils with inadequate data would have been zero-gain or low-gain pupils, anyway. To local district administrators struggling with already tight budgets, the performance-based funding provisions of Chapter 3 must have appeared intolerable, and they demanded that enforcement be put off again (Murphy and Cohen, 1974).

The MDE -- aside from the question of whether it could have resisted the pressure -- saw some justice in this demand. As one staff member said,

> ... you still have kids there to serve. What we're interested in, as educators, is serving those kids with the best possible programs. By withdrawing money, we're running counter to doing that.

This time, however, instead of starting still another three-year cycle -- which might have been too embarrassing and might have meant the demise of the TBF program altogether -- a different tactic was adopted, in the form of a new provision, Section 39a, which was added to Chapter 3 in the summer of 1973. Districts which, because of their 1972-73 test results, would fail to "earn" for 1973-74 a portion of their original allocation could reapply for these funds

by proposing "a different educational delivery system than was provided for students who did not achieve 75% of prescribed minimum performance objectives in 1972-73." Most of the districts did reapply and received virtually their entire original allocations for use in 1973-74.[6] The MDE issued guidelines for the reapplications, urging some form of individualized instruction, but this was largely ritualistic. Actually, it is difficult to see how a new delivery system could have been proposed specifically for those pupils who had not achieved the 75 percent gain, since they were not clustered together but rather were scattered as individuals among the school buildings and classrooms of the district. In practice, the districts' proposals were quite perfunctory,[7] and the MDE did not insist on any changes. A Department staff member described the process this way:

> Districts filed assurances that said, basically, they would follow the guidelines. Later in the year, we picked up

---

[6]Section 39a withheld some funds from the reallocation process: up to two percent "for the employment of an external and independent agency for monitoring the contractual arrangements [presumably meaning the proposed new delivery systems] and validating the results thereof," and up to $100,000 for "a cost-effectiveness study of Michigan compensatory education programs." These deductions were prorated among the districts. In addition, the MDE ruled that, while a district could reearn funds lost on account of pupils for whom scores "were reported, but were inaccurate," it could not reearn funds lost on account of pupils for whom no scores were reported at all (1972-73: [Part Two], p. 33). Finally, 17 small districts with only a few Section 3 pupils who had not made the 75 percent gain did not reapply. In the end, $4,241,000 were reallocated for 1973-74 under the provisions of Section 39a -- all but about $550,000 of what otherwise would have been lost.

[7]Murphy and Cohen (1974) quote these examples: "Changed grouping practices. Lowering class sizes. Providing Reading-Readiness." "Program to strengthen discovered areas of need, and prescribe treatment." "Individualized instruction and remediation in reading and math."

> descriptions, after they were operational. In the meantime, we hired an independent auditing firm to go out and audit for compliance with the guidelines .... On the basis of the auditor's report, we considered we had excellent compliance.

In effect, then, the performance-based funding provisions were waived again for 1973-74.

When the 1972-73 test results were published, in February, 1974 (1972-73: Part One), they showed substantial improvement over the previous year (Table 2.2). The number of pupils for whom adequate data were not available had been cut in half. Among those for whom scores were reported, the proportion making no gains had dropped to nine percent, from the 16 percent of the preceding year, and the proportion making partial gains had gone up to 31 percent, from 18 percent. On the other hand, the proportion making 75 percent gain or more had gone down, from 66 percent to 59 percent. That decline was concentrated in Detroit (67 percent to 50 percent). Had it not been for Section 39a, the 67 districts would have lost a total of nearly $5,000,000 — less than the $8,000,000 of the preceding year, but still a substantial sum of money, and $3,000,000 of it would have been lost in Detroit alone. With little dispute, Section 39a was renewed for 1974-75.

Efforts at evaluation

It is impossible to say whether Chapter 3 has been "successful." For one thing, there is a question as to what criterion of "success" is to be used. The MDE apparently hoped that Chapter 3 would have two direct effects: it would improve the performance of low-achieving

Table 2.2. Potential 1973-74 Allocations under Michigan's Chapter 3 Program

	Detroit		Other 66 districts		Total for 67 districts	
	Number of pupils	Potential allocation	Number of pupils	Potential allocation	Number of pupils	Potential allocation
Total for 1972-73	59,267	$11,853,400	53,233	$10,646,600	112,500	$22,500,000
Pupil Category for 1973-74 funding						
Made gain of 75% or more	25,029	$ 5,005,800	33,733	$ 6,746,600	58,762	$11,752,400
Made gain of 1-74%	19,273	$ 2,415,104	11,895	$ 1,499,228	31,168	$ 3,914,332
Special eligibility	6,454	$ 1,290,800	3,781	$ 756,200	10,235	$ 2,047,000
Subtotal	50,756	$ 8,711,704	49,409	$ 9,002,028	100,165	$17,713,132
% of 1972-73	86	73	93	85	89	79
Made no gain	5,688	($ 1,137,600)	3,430	($ 686,000)	9,118	($ 1,823,600)
Data not available	2,823	($ 564,600)	394	($ 78,800)	3,217	($ 643,400

Source: 1972-73, Part One, p. 8.

pupils, and it would increase efficiency in the use of public funds. The Department did attempt to demonstrate that the performance of funded pupils was indeed improving more than might otherwise be expected. For instance, in the final report on the 1971-72 year, data for Lansing were presented showing that, although its Section 3 pupils had begun the year at 0.2 to 1.4 grade-equivalent years lower in reading and mathematics than pupils in the corresponding grades in the city as a whole (thus implying that they had been progressing at a lower rate previously), the Section 3 pupils gained almost as much during the year as did pupils in the city as a whole (1971-72: Report Number 2, pp. 34-35). But there is no indication of how representative Lansing is of the 67 districts, or of why it was chosen for this comparison, nor is any mention made of the fact that scores were missing for 16 percent of Lansing's Section 3 pupils (1971-72: Report Number 2, p. 112). Furthermore, a measurement of grade-equivalent gain taken between fall and spring usually will give the impression of greater progress than a measurement of the average grade-equivalent gain made over several years, merely because grade-equivalent scores are based on the assumption of smooth linear gain from the testing point in one year to the testing point a year later, ignoring the losses that typically occur during the summer (Horst and others, 1975, p. 31). Thus, these figures cannot be generalized into a conclusion about the effectiveness of the Chapter's provisions.

The next year, 1972-73 pretest data for a large number of districts were used to estimate the grade-equivalent scores that could be expected on the posttests, according to the test publishers' norms; since the actual posttest scores of the Chapter 3 pupils were about half a year higher than these expected scores, in both reading and mathematics in all grades (2-6) for which the calculations were made, the MDE concluded that "children receiving additional educational services provided by Chapter 3 monies achieved at a higher rate than could have been expected had they been in regular school programs" (1972-73: [Part Two], p. 21). But there are several defects in this analysis, and so it is not altogether persuasive, either.[8]

Even if Chapter 3 pupils did achieve more than expected, the explanation — and thus the lesson to be inferred — would be uncertain. It may have been a result simply of the additional funds that districts received, which may have allowed them to reduce the pupil/staff ratio or to develop or install new programs

---

[8]The major problem lies in the fact that the expected scores used in the comparison were a simple unweighted mean of the expected scores for four tests, derived from the respective test publishers' tables for converting percentile ranks into grade-equivalent scores (1972-73: [Part Two], pp. 56-65). On the one hand, it is extremely unlikely that each of the four tests was administered to just one-quarter of the pupils included in the analysis; and on the other hand, the conversion from percentile ranks to grade-equivalent scores varies considerably from one test to another. The differences between the actual and the expected scores may therefore reflect, at least in part, merely the differences among the conversion tables combined with the distribution of the pupils among the tests. The number of districts included in the analysis is variously given as 66 and as "approximately 50," and there are other obscurities in the procedure as well.

of compensatory education.[9] This is what the Department seemed to suggest in its conclusion from the analysis of 1972-73 data that has just been quoted; but if that was the case, the provisions for the following years, reducing funds to the degree that pupils did not achieve the required gain, would have been counter-productive.

Perhaps, however, it was the result of those very provisions, which may have impelled districts to strive to improve instruction in order to avoid the loss of funds. It is true that, because of the postponement for 1972-73 and then the Section 39a reallocations, few districts ever really lost funds, yet that does not necessarily mean that the incentive was not operating. Each year, administrators and teachers may have thought that funds would be lost if pupils did not make the required gains and they may have acted accordingly.

A difficulty with that line of reasoning is knowing just what "acted accordingly" would mean. What did administrators or teachers do that they would not have done in the absence of the fear of losing funds? As we have mentioned before, Chapter 3 pupils were not grouped together in any way but were scattered through many of the buildings and classrooms of a district; thus, it is a virtual certainty that they were not the distinctive targets of any special instructional programs. Most teachers probably did not know which

---

[9]The Department's analysis of expenditures did not find any pattern of spending that sharply differentiated districts with large proportions of high-achieving pupils (those making at least the 75 percent gain) from those with small proportions; the biggest difference was that, in the second year, high-achieving districts spent $13 more per pupil (of their Chapter 3 funds only) for instructional salaries than did low-achieving districts (1971-72: Report Number 2, pp. 47-55; 1972-73: [Part Two], pp. 25-29).

of their pupils were the Chapter 3 participants; and if they had known, it seems very unlikely that they would have "tried harder" with those pupils than with others in the same classroom. Perhaps their efforts led to improvement in the performance of all K-6 pupils in the district, but the type of analysis required to detect that would have been much more complicated than anything attempted by the MDE (or anything that could be attempted within the scope of this study). In short, no firm statements can be made about the effect of Chapter 3 on the level of pupil achievements.

With respect to the efficiency aspect of Chapter 3, MDE staff members made statements such as these:

> A district that isn't successful [in improving student performance] could be judged to have had an opportunity to serve the children and to be successful. Since we have students with the same needs in other districts, then as a public policy, it may be a better choice to move those monies from that district to another one and give them that same opportunity to mount effective programs with other kids.
>
> You always have kids that are in need that aren't being served. But school districts are in the public policy-making arena, so there's every reason to say to them, "If you cannot show success with that $200, why should we give you $200 to show us that you can't succeed again?"

However, it had never been anticipated that Chapter 3 money unearned by one district would be redistributed to another. Indeed, there was a cogent reason why a district that had been unsuccessful with its initial $200-per-pupil allocation should get it back again, anyway: "You still have kids there to serve." If one of the basic

arguments that had been used in support of TBF was to be upheld -- that test scores were a valid measure of educational need -- then there would seem to be no choice but to continue funding districts for their low-scoring pupils.

Operational problems and future prospects

A number of problems arose at the operational level of Michigan's TBF program which involved, in one way or another, the tests, the testing procedures, and the measurement of gain.

As we have pointed out, most of the pupils who were supposed to be the beneficiaries of Chapter 3 funding were selected from among those who had scored lowest on a standardized achievement test, and a district's funding in the second and subsequent years was to be based on the extent to which the scores of these pupils increased over the course of the year. However, because every test is less than perfectly reliable, the low scores on the selection test would be partly due to measurement error; if the selected pupils were to be given a second test, it is unlikely that the error would be repeated in exactly the same way, and so their scores on the second test would tend to be higher than those on the first test even if no learning had occurred in the interim. This phenomenon is known as "regression toward the mean." Its import in the present context is that, if the selection test were also used as the pretest -- i.e., the baseline from which achievement gain

was determined -- the posttest scores would show greater gains than had "truly" taken place and a district would receive more funds than it would if gains had been measured without error.

To avoid this, the MDE required that, after pupils had been selected, they were to be administered another test which would serve as the pretest. Thus, "whatever regression there is will affect the relationship between the test [used] to determine eligibility and the pretest and not [the relationship between] the pretest and the posttest" (1972-73: [Part Two], p. 17).[10] While this may have had the desired effect, the regression phenomenon nevertheless cropped up in another and somewhat discomfiting way. The Department found (1971-72: Report Number 2, pp. 63-65) that about 30 percent of Chapter 3 pupils were "misplaced" (1971-72: Report Number 1, p. 26) in the sense that, on their pretests, they failed to meet the eligibility criterion of scoring one year or more below grade level. Some of this may have been due to errors in teacher judgments where they were used to determine eligibility, and some to the inability of instruments to yield scores one year below grade level for kindergarten and first-grade pupils; but a large part of the difference was surely attributable to the regression that had deliberately been "allowed" between the selection test and the pretest. Thus, a statistical trap had been avoided only at the expense of raising a question about the degree to which Chapter 3 was serving the pupils it was supposed to serve

[10]This was a technique suggested by Wrightstone, Hogan, and Abbott (no date).

(an unanswerable question, since there is no reason in principle to prefer the results of either the selection test or the pretest over the results of the other) -- and, it may be added, at the more literal expense of an extra test administration.

Another problem was the selection of the instrument to be used for measurement of gain. Because Chapter 3 pupils were low achievers, some school districts gave them a pretest that had been designed for pupils at a lower grade than the one they were actually in -- e.g., a second-grade test for third-graders -- in order that the test not be too difficult for them. The result, though, was that the test was then too easy for some pupils; they scored at or near the maximum and were thus unable to show any gain on the posttest (1971-72: Report Number 1, pp. 17-18). In accordance with state policy, the Department was encouraging districts to move toward the use of criterion-referenced tests instead, and it may have hoped that this would eventually resolve the problem by allowing objectives and test items to be chosen to suit the performance level of each pupil, but it is not clear that this hope would be vindicated. Some districts had used criterion-referenced tests even in the first year, and the Department itself had complained of the lack of "guidelines [for] dealing with" such tests (1971-72: Report Number 1, p. 14), which probably meant above all the difficulty of deciding what a "suitable" objective was.

The tests were administered, and the scores reported, by local school-district personnel, and this, too, created problems. Publishers' instructions for test administration were not always rigorously followed; students were not always uniquely identified in the score reports, so that pretest and posttest scores sometimes could not be matched; and some districts did not report scores at all but rather letter grades or verbal descriptions such as "high," "average," or "low" (1971-72: Report Number 1, pp. 17-19). The demands on the districts were of course magnified by the fact that scores had to be reported twice a year for individually identified pupils who sometimes numbered in the thousands. As would be expected, testing and reporting went more smoothly in the second year, but even then, improvements may have been limited because of a 31 percent turnover rate among the local Chapter 3 contact persons between the first and second years (1971-72: Report Number 1, p. 20). All of this was a price paid for the economy, and perhaps the politically more acceptable tactic, of local rather than state test administration.

A complaint made by school administrators was that, once the initial allocations had been made, a district could not increase the amount of funds it received under the program. There was, for example, no extra money for a district in recognition of pupils who may have made more than the required gain. Thus, from the district's point of view, Chapter 3 was a merely punitive device; it could lead to a reduction of aid funds but not to an increase.

The MDE, and the State Board of Education, have come to recognize this themselves. In the process of shaping their position on the future of state aid for compensatory education, they proposed that $150 of the original per-pupil allocation be converted into a permanent part of state aid (i.e., no longer based on pupil scores, except in an historical sense) and that the other $50 be used as a genuine incentive, to be awarded to districts whose pupils improved in achievement. It should be noted that this would imply abandonment of the concept that low test scores indicate the existence of educational needs. The proposal was rejected by the legislature, apparently because it would still have meant, for many districts, a reduction of state aid below the present $200-per-pupil allocation.

Chapter 3 was scheduled to expire at the end of the 1974-75 school year. As this is being written, the legislature has not yet taken final action on a renewal, but the most likely outcome of the deliberations appears to be that TBF will be continued in some form, though almost certainly without provision for required gain in order to maintain the original allocations. Instead, as a substitute method for dealing with the disincentive problem, districts will probably have to obtain MDE approval for their compensatory programs. The total appropriation will apparently remain at $22,500,000 for the fourth successive year, but the funds may be

allocated afresh on the basis of a more recent administration of the state assessment battery.[11]

California[12]

California has two programs with TBF provisions. They differ from Michigan's Chapter 3 program in an important respect: The TBF provisions have not played nearly so central a part in them. Consequently, the disincentive problem has been of lesser magnitude and it has been dealt with not so much as a problem in its own right but rather as a byproduct of a drive for greater state control over local districts' use of funds.

The first of these programs was the Miller-Unruh Basic Reading Act of 1965. This act was intended, among other things, to induce school districts to employ specialist teachers in reading for the early grades. Each district was to receive annually from the state an amount of funds for this purpose which was equal to $250 more than the average salary of elementary-school teachers in the state, multiplied by the number of specialists it actually employed up to the quota for the district. The quota was set at one specialist for each 125 pupils in average daily attendance (ADA) in grades 1-3,

---

[11]The tests of the battery were converted into criterion-referenced instruments in 1973. It is not known how, or whether, this would affect the allocations of Chapter 3 funds. For an interesting discussion of other consequences of the change, see Murphy and Cohen (1974).

[12]The following account is based on interviews with several staff members of the California Department of Education and on the departmental publications cited below.

except that -- and this was the TBF provision -- in districts where 30 percent or more of the first-grade pupils received scores below the first quartile on a statewide reading test, the quota was increased by one specialist for each 300 pupils. The disincentive issue was not raised, but it may have been presumed that, since a district could use these funds only for paying the specialists' salaries, it would have little to gain by depressing or manipulating test scores.

The appropriations for the act, however, have never been adequate to pay for all the specialists who could have been employed under its provisions. Indeed, they have not even been sufficient to cover the salaries of the specialists that districts were entitled to hire under their basic ADA quota. The impact of this circumstance has been aggravated by other provisions of the law. The specialists had to meet certain criteria of certification, and a district had to hire them before it could apply for the funds. Thus, the districts that were able to qualify earliest were those that already had specialists in their employ who met the certification requirements, and these tended to be relatively well-to-do districts with small proportions of low-scoring pupils. Yet the law also stated that, once a district qualified, its Miller-Unruh funds could not subsequently be reduced as long as it retained the specialists in its employ. The net result was that funds were never available for specialists under the part of the quota based on test scores, and so the TBF provisions have remained a dead letter.

The other California statute which led to a TBF arrangement is SB1302, the Early Childhood Education (ECE) act, which became operational in September, 1973. It authorizes grants to school districts which "develop and submit to the Department of Education for approval a master plan" for upgrading the quality of educational programs in kindergarten through the third grade. ECE, it should be emphasized, is not a compensatory-education program. California does have a state-level IBF compensatory-education program, SB90 (Education for Disadvantaged Youth), and ECE has not replaced it or other forms of state aid. The appropriations for ECE were $25,000,000 in its first year and $40,000,000 in its second year, compared to more than $80,000,000 for SB90. Furthermore, any district can qualify for ECE funds, without regard to its socioeconomic composition or the performance of its pupils, and in fact about 90 percent of the districts in the state have qualified. Taken together, these attributes -- together with others to be noted below -- greatly relieve the pressure on the TBF aspect of ECE.

ECE grants, like those in Miller-Unruh, are divided into two parts: $130 (later raised to $140) per pupil in ADA in the participating schools and grades, and an additional $65 (later raised to $70) for each pupil "determined ... to have demonstrated educational need." The criteria of "educational need" were given in the law as "low levels of pupil achievement and such factors as low levels of family income," but were otherwise left to the

State Board, Superintendent, and Department of Education to specify; and they, in practice, have defined a pupil with educational need exclusively as being one who scores in the lowest quartile of a statewide test in reading and mathematics.

This choice was a matter of expediency, not of principle. The proposal for an early childhood education program which was presented by the California Department of Education (CDE) to the State Board early in 1972, and which was the basis for the subsequent legislation, did not even mention test scores. It did call for the $130 and $65 per-pupil grants, but suggested that the latter go to districts according to the number of children from low-income or "disadvantaged" families (California State Department of Education, 1973a, pp. 10-11). After SB1302 had been passed, the CDE's own guidelines for implementation said, rather ambiguously, that educational need "shall be defined as including those pupils in the lowest quartile in reading and mathematics according to state achievement tests" or those who qualified under the terms of SB90 "relative to 'potential impact of bilingual-bicultural pupils,' 'index of family poverty,' and 'index of pupil transiency' ..." (California State Department of Education, 1973b, p. 9). But statewide tests had been administered annually to all pupils in the first three grades ever since they had been mandated for purposes of Miller-Unruh. Thus, the test-score data were readily available, and using them gave a new justification for the state testing program which, as we have seen, was never

actually used for Miller-Unruh.[13] They were also, as one CDE staff member put it, "cleaner" than measures of poverty or transiency. Even at that -- and underscoring the fact that ECE was not a compensatory-education program -- the $65-per-pupil allowances were limited to 25 percent of the K-3 pupils in any one school, which meant that they could account for a maximum of only 11 percent of a school's total ECE funds.[14]

As this history indicates, neither the legislature nor the education agencies in the executive branch have regarded TBF as crucial to ECE, nor was the appeal of TBF responsible for its passage or the size of its appropriations. The salient component of the act, rather, was the accountability which was built into it and which was to be enforced by the CDE. However, for purposes of the present study, this accountability can be viewed as a safeguard against the potential disincentive effects of TBF, and an examination of how it has worked is therefore in order.

As has been mentioned, a district qualifies for ECE funds only after the CDE has approved its master plan for an early childhood

---

[13] A similar circumstance may have been part of the impetus for Chapter 3 in Michigan; see Murphy and Cohen (1974).

[14] If a school's K-3 ADA is X, then the maximum proportion of its ECE funds which can be derived from its grants for pupils with educational need is $\frac{\$65(.25X)}{\$130X + \$65(.25X)} = .11$.

education program. The plan is to be implemented, in its first year, in only some of the K-3 grades and/or some of the schools in the district, and it is to show how the program will be expanded to the rest of the grades or schools over a five-year period.[15] Each school seeking participation also submits its own plan; originally, these plans were reviewed and rated by the CDE, and only the highest-rated schools in qualifying districts were permitted to participate.

The law requires that, before a district can be granted funds for expanding the program into other grades or schools, the CDE must determine that the district has "met the objectives of its approved plan" for the previous year. Failing such a determination, funding for the district and its participating schools was not to be decreased but would simply be continued at the same level. These provisions constitute a stimulus for both districts and schools to put forth their best efforts and to avoid artificially depressing test scores in order to increase their funding -- a kind of "functional equivalent" to the requirement for test-score increases that was a feature of Michigan's Chapter 3, yet without leading to the necessity for a reduction in state aid. Out of concern for the effective use of funds over the long run, though, the law did order the State Board of Education to "adopt rules and regulations governing the termination of allowances to districts which are unsuccessful in meeting the objectives of their approved plan." The rule adopted

---

[15]The ECE Act stated that, in each year prior to the last, at least half a district's ECE funds had to be spent in those of its schools with the largest numbers of pupils having "educational need." The CDE has interpreted this to mean that at least half of the participating schools in the district must be among those with the largest numbers of pupils scoring in the lowest quartile by state norms. Some participating districts have relatively few such pupils.

was that a district would be "considered for termination" if it was refused expansion funds for two successive years.

These provisions generated a heavy workload at both the local and state levels. First, each district and school had to draw up its plan. Proceeding from the general authorizations of the act, the Department required (California Department of Education, 1973b, pp. 3-6) that the plan include objectives for pupil performance in reading, language, mathematics, and "other curricular areas"; for the reduction of pupil/adult ratios (through the use of aides and parent volunteers) and other steps toward individualized instruction; for bilingual education where appropriate; for pupil health; for staff development; and for parental education, and parental involvement in the planning, implementation, evaluation, and modification of the program. These objectives were to "be stated in language that is concrete, unambiguous, and capable of measurement or observation." So formidable a set of requirements may have deterred application from some districts that lacked the capability for preparing such rigorous plans, or perhaps the CDE found the careful review of a large number of detailed plans beyond its capacity; in any event, approval was given to the plans of all 800 districts that applied. They involved 1,010 schools enrolling 14 percent of the state's K-3 pupils (California State Department of Education, 1974, pp. 6, 11).

During 1973-74, the CDE undertook to determine how well each district had met the objectives in its plan. For this purpose, the

Department developed an elaborate "school-level quality review" procedure, under which one or more persons would visit a school and rate each of more than 60 elements of program and management on a ten-point scale. Each school was given a score consisting of these ratings, the ratings on two self-evaluations which it had to submit during the year, and the rating on the plan by which the school had qualified for participation in ECE. The school scores were then averaged to yield a district score. The district scores were arrayed from high to low, and the districts with scores among the highest 80 percent were informed that they were eligible for expansion to an additional proportion of their K-3 ADA ranging from 10 to 60 percent, according to their standing in the array and subject to legislative appropriations. At that point, the newly entering schools had to submit their plans for approval -- and some districts discovered that, although they had authorization to expand, none of the schools in which expansion was to take place was able to present an acceptable plan. Districts with the lowest 20 percent of the scores were refused permission to expand at all; most of them were small districts, but they also included San Francisco and Berkeley.

The burden of these procedures proved to be too much for the CDE to handle; the on-site visits alone occupied 22 staff members full-time, and a number of part-time consultants, over a period of more than three months (and 140 small schools were not visited). In the fall of 1974, several changes were made. First, any district

not yet participating would, at its request, receive authorization for a K-3 program at one school in 1975-76, and any school with an ECE-funded program operating in only some of grades K-3 would automatically be authorized to expand its program to all four grades. Secondly, the requirement was dropped that school plans be approved for newly entering schools in an expansion district; instead, expansion was to take place in accordance with the district's master plan (although the schools still had to submit their plans in order to have objectives whose attainment could later be observed). Thirdly, the on-site visits were limited to schools with the lowest 1973-74 scores and those entering the program for the first time in 1974-75. For schools that were not visited, the rating they had received in the previous year was used again in calculating the district score.

In 1974-75, decisions had to be made about expansions for 1975-76. These decisions were based on the same information as was used in the previous year, with one addition: Each school that had participated in 1973-74 was to make a report on the attainment of its pupil-performance objectives -- i.e., test-score gains -- during that year. Interestingly enough, the ECE act stipulated that the degree of attainment of these objectives was to be given a weight of only 10 percent in a school's score in the first year of its participation, though the weight was to rise to 40 percent in its second year and 50 percent in its third. Reference to test-score

gains raises the TBF dilemma -- that if a school managed to improve its educational program to the point where test scores began to increase, its funding would decrease insofar as the amount had been based on low scores. This apparently has not yet happened in any school, and its impact would be limited because of the small proportion of funding derived from this source. Nevertheless, the CDE is aware of the possibility and is considering what might be done. The two proposals receiving most active consideration are that a school's proportion of pupils in the lowest quartile at the time it entered ECE become a permanent part of its allocation, whatever happened to the proportion afterward; and that the amount of the allocation simply be increased for all schools by one-eighth (i.e., the one-fourth of the pupils eligible for the low-score allocation multiplied by the one-half of the basic per-pupil allocation which was made for those pupils). Either of these steps would mean, in effect, the elimination of the TBF provisions.

District scores for 1974-75 were once again arrayed, and districts with the lowest 20 percent of scores were denied expansion funds. However, only 10 districts were refused authorization for expansion in both years; all of them were small districts, neither San Francisco nor Berkeley being among them. It is the CDE's intention to give these districts "intensive attention" in 1975-76 so that their participation will not have to be terminated.

The usual claims of success have been made for ECE, but the CDE's evaluations do not clearly bear them out, at least so far as pupil gains in reading and mathematics are concerned. The first evaluation report (California State Department of Education, 1974) made comparisons only among pupils in various categories of ECE schools -- those receiving ECE funds only and those receiving both ECE funds and funds from other special-aid sources. When the complaint was made that these comparisons did not show how pupils in ECE schools had fared relative to pupils in non-ECE schools, the CDE produced a supplementary report ([California State Department of Education], 1975) which made much of the fact that pupils in ECE schools had fairly regularly made gains that were very slightly greater than those made by pupils in non-ECE schools. But the comparisons are rendered virtually meaningless by a host of complicating factors, including particularly the heterogeneous SES composition of ECE pupils and the fact that districts and schools had been selected for participation in the ECE program by a process that suggested they may have had superior instructional capabilities to begin with. It is also pertinent to repeat that no participating school had had its funds reduced on account of a decline in the proportion of its pupils in the lowest quartile.[16] On the other

---

[16]The absence of such a decline is not incompatible with the gains among pupils in ECE schools mentioned above. The former refers to different pupils in a given grade, one year apart (e.g., third-grade pupils in 1973-74 and third-grade pupils in 1974-75); the latter, to the same pupils tested at the beginning and end of the school year.

hand, the Superintendent of Public Instruction, Wilson Riles, has asserted, with some justification, that the success of such a large-scale "restructuring" as ECE is supposed to bring about cannot be judged from the test scores of pupils in its first year.

What can be said is that the TBF provisions of ECE have created far less severe problems than have those of Michigan's Chapter 3. In large part, this is because they have played a rather minor role in determining the funding for a school or district -- and even at that, there is no serious threat that the funds will be reduced or terminated. On the contrary: there is an opportunity to obtain additional funds for expansion, and it is not heavily dependent on pupil test scores, either. It is reasonable to believe that this opportunity acts as an incentive to local administrators and teachers, not merely because of the financial aspect but also -- and perhaps primarily -- because it represents a symbolic recognition of the worth of their efforts. One department staff member reported that there was "a lot of hurt pride" in San Francisco when it was denied expansion funds at the end of the first year, and administrators and teachers "worked very hard" to qualify for expansion in the next year.

Meanwhile, the demands growing out of the accountability requirements have been somewhat reduced, as we have seen, and perhaps they were not very great in the first place. A Department staff member said that one reason the competitive rating of school

plans was dropped was that there was so little difference among them, anyway. Even the extensive monitoring activities of the CDE appear to have provoked little objection. Some "sensitivity" to the on-site visits was reported during the first year, but the CDE itself found it necessary to cut back on the number of these visits because it could not handle the workload. In effect, the accountability procedures have become a form of technical assistance to local districts and schools, with a carrot if the assistance is used well and a small stick if it is not.

New York[17]

The TBF programs in New York have been less troublesome than those in either Michigan or California. The main reason seems to be that few strings have been attached; the programs have been devices for funneling additional aid to districts believed to need it, with no pretense of promoting dramatic breakthroughs in educational reform.

In 1967, the state established an Urban Education program. Under its provisions, a district was eligible for supplementary funds if (1) it had a weighted ADA (WADA)[18] of more than 4500 pupils in

---

[17]The description which follows is chiefly based on interviews with several staff members of the New York State Department of Education.

[18]New York counts each pupil in grades 7-12 at 1.25, in consideration of the greater costs of secondary education.

all grades, (2) at least five percent of its school-age pupils (in both public and nonpublic schools) were from families receiving AFDC assistance, and (3) it had a Special Needs Index (SNI) of at least 1100. The SNI was the district's WADA multiplied by the percentage of sixth-grade pupils who had scored below the fourth stanine in reading on the 1967 administration of the PEP test. By these criteria, 29 of the state's 750 districts qualified, and the amount of the grant for each was a share of the total appropriations for the program, which were $46,300,000 in 1972-73. A district's share was based on its proportion of sixth-grade pupils scoring below the fourth stanine in 1967 among all such pupils in the state, and thus remained constant throughout the life of the program (New York City received nearly 85 percent of the funds). The PEP test is administered annually by classroom teachers, and according to one staff member of the New York Department of Education (NYDE), "there isn't too much control over the testing conditions." Use of scores from the 1967 administration, however, precluded manipulation subsequent to enactment of the legislation.

Districts were expected to propose specific projects to be supported by their Urban Education funds, designed to facilitate their compensatory-education efforts, and these projects were subject to approval by the NYDE. But, one staff member said, "We weren't always able to check them out before approval or to follow them up later." Indeed, in most cases projects did not receive final approval until statements of expenditures were submitted at the end of the

year; expenditures violating the guidelines could then be disallowed (and were to be deducted from the next year's funding), but this evidently did not happen very often (Berke and Kirst, 1972, p. 353). Projects initiated in one year could be continued in subsequent years with even more cursory review, and in 1972-73, more than half the projects had been in existence for three years or more.

Data provided by the districts for about one-third of the pupils participating in Urban Education projects in 1972-73 showed mean gains of about one month in grade-equivalent units for each month of instruction, in each of four curricular areas -- reading, mathematics, "other basic academic skills," and "other than basic academic skills" (University of the State of New York, 1974, p. 86). However, as the evaluation report pointed out, there were no "comparable control groups," and many if not most of the Urban Education projects also received funds from other sources, federal or state or both.

In 1974, Urban Education was repealed and was replaced by a program called Pupils with Special Educational Needs (PSEN). TBF was retained, but in a somewhat different form, adapted from the Fleischmann Commission's recommendations (see above, p. 18). The eligibility criteria were dropped, so that nearly all districts receive some PSEN funds. The basic funding provision is that, in determining a district's aid under the state's general aid formula, a proportion of its ADA is given a weight of 1.25, the proportion

being equal to the average of the percentages of its sixth-graders with a composite score in reading and mathematics below the fourth stanine of the PEP tests administered in 1971 and 1972. Use of the 1971 and 1972 scores allowed the distribution of funds to reflect more nearly current conditions than would continued use of the 1967 scores that had been the basis for Urban Education aid; yet they were still pre-enactment scores, a decision deliberately taken, according to one staff member, both to avoid the possibility of manipulation and because "we didn't want to penalize districts that improved." On the other hand, the fact that they were pre-enactment scores admittedly meant that they would eventually suffer the same obsolescence as the 1967 scores. The percentages in two years were averaged to reduce the effects on a district of erratic year-to-year variations in the scores.

The NYDE regarded PSEN simply as a way of getting additonal money to all districts that needed it, rather than only to urban districts. Measures of need other than test scores were considered but rejected: AFDC because it was believed to be affected in rural areas by seasonal employment patterns and "a more independent feeling" on the part of residents in those districts; numbers of pupils in school-lunch programs because "some districts keep closer watch on the program than others," and a few do not participate in it at all; and family income because the data were "suspect and difficult to use." But it was assumed that test scores were as valid a measure

of educational need as any of these SES characteristics, and they were administratively more convenient and readily available and had been used for some years in the Urban Education program without repercussions. Thus, as in California, the choice of test scores as the measure of need was a matter of expediency rather than of principle.

The proper grade level for the tests was a subject of some disagreement. The Fleischmann Commission had urged that eventually a "learning-readiness" test be adopted that would be given to pupils as soon as they entered school; the Commission believed this to be the surest way of avoiding a disincentive effect.[19] Acknowledging that a reliable test was not yet available at that level, the Commission recommended the use of a third-grade test as an interim step that would at least minimize the disincentive and yet allow an immediate start on (or expansion of) TBF, the third grade being the earliest one at which New York already administered a statewide test (Fleischmann Report, vol. 1, p. 68). The NYDE, however, saw matters in another light. Differences in instructional programs in the early school years might have an effect on student performance at the third grade. For example, with respect to children whose native language is not English, some schools may start out by teaching only in English: others may start by teaching students in their native

---

[19]For further discussion of this strategy, see below, pp. 71-73.

language, gradually shifting to or adding instruction in English; still others may teach in English and another language simultaneously, with resultant variations in the pupils' comprehension of written English at the third grade. Similarly, in mathematics, some schools may emphasize computations in the early grades while others stress the understanding of concepts. The NYDE, a staff member pointed out, had no wish to impose any one of these approaches on all districts, nor any reason to believe that one of them was consistently superior to the others.

Consequently, on the ground of "fairness" to the districts (i.e., preservation of local autonomy), the NYDE prevailed upon the legislature to continue using the sixth-grade test instead, for by that time the effects of these early differences ought to have disappeared and it was reasonable to expect that all students should be on a common level of learning if they had been taught equally well by whatever method their school had chosen. Another way of interpreting the NYDE's view is to say that it was less concerned with the possibility of a disincentive than it was with the possibility of manipulating scores to produce the appearance of low performance. If the third-grade test were to be used, schools might be inclined to adopt methods which produced the lowest scores at that point, especially if they believed that it would make no difference to their students' learning in the long run.

Under the PSEN program, districts know how much money they will receive each year, and they are expected to plan projects accordingly and notify the NYDE, but project approval essentially occurs at the end of the year when expenditure statements are submitted. Detailed guidelines for the use of PSEN funds have not yet been issued, but they will apparently be similar to those used for Urban Education. It is estimated that 1974-75 PSEN projects will be supported by about $75,000,000 of state aid. Districts are supposed to designate target pupils, but the NYDE has defined these pupils as being those who, in the judgment of local administrators and teachers, have "any indication that they have serious learning deficiencies." An interesting proviso of PSEN, whose implications will be explored in the next chapter, is that up to 30 percent of the funds may be spent on otherwise ineligible pupils if that is necessary in order to avoid segregating target pupils in special classes. The program is, of course, too young to permit any statement about its effectiveness in improving pupil performance.

Summary and inferences

This exploration of the experience of three states with TBF programs does not yield conclusive findings, if only because it relies largely on the view from the respective state departments of education. Moreover, it is impossible to say whether this experience would be duplicated in other states with different characteristics or in a TBF program at the national level. Still, it is the only concrete

experience we have, and we would be remiss if we did not try to draw some tentative inferences from it. Broadly speaking, what it suggests is that the seeming advantages of TBF are more difficult to realize than might have been thought, and indeed that they may be closely linked to its drawbacks.

1. One benefit of TBF is supposed to be that it allows the data base for funding to be kept more up-to-date than the Census figures on income can be. But this requires that the test used to determine allocations be administered at fairly frequent intervals, and if that were to be done, it would create opportunities to depress or manipulate the scores for the sake of increased funding. To prevent that, both Michigan and New York based their TBF allocations on scores on tests administered prior to enactment of the legislation and continued to use the same scores for several years. The scores were updated only when the legislation was renewed or revised. If this updating becomes regularized, the temptation to manipulate is recreated; if it is not, the data become obsolete.

2. TBF is supposed to lead to a more efficient (less wasteful) distribution of funds by channeling them directly to where they are most "needed" -- i.e., to low-scoring pupils. Such precise targeting, however, is not easily accomplished. All three states found it administratively awkward or pedagogically undesirable (or both) to segregate low-scoring from other pupils; New York's PSEN program, in fact, explicitly permits up to 30 percent of the funds to be

spent on non-targeted pupils where necessary to avoid such segregation, and the ECE program in California is not aimed primarily at low-scoring pupils at all. In the absence of this type of segregation, the same sort of "resource leakage" occurs which Benson spoke of as being one of the disadvantages of IBF (see above, p. 10).

3. TBF is also supposed to make state aid more effective in improving pupil performance by tying the aid directly to low scores or to "accountability for results." There is as yet no clear evidence that scores have risen in any of the three states, though PSEN and ECE are still too new to expect such evidence to appear. Meanwhile, a dilemma has emerged in the implementation of TBF: What is to be done if test scores do not increase? To withdraw or reduce the test-based funds either amounts to a refusal to supply aid where it is needed or belies the premise that low scores are an indication of need; to maintain the aid at the same level allows the ineffective expenditure of funds to be prolonged. Michigan found it politically impossible to carry out the threat contained in the Chapter 3 provisions that aid would be reduced to the extent that scores did not improve. California's ECE program includes a provision for termination; the state has so far not had to try to put it into effect but is concerned that it may have to and is considering changes in the law to avert the necessity. New York has "solved" the problem by ignoring it; no demands for strict accountability are made on the local districts, and no sanctions are imposed in the event that test scores do not rise.

4. The ability of TBF to produce larger appropriations of funds for education is doubtful, at least at the state level. In Michigan, it led to a one-time increase to a level which then remained constant for at least four years (and constant in actual dollars, which in effect meant a decline in the purchasing power of the funds). California's first attempt at TBF, the Miller-Unruh Basic Reading Act, never was funded at a level high enough to allow the TBF provisions to take effect; its second attempt, the ECE program, was accompanied by increases in state aid during its first (and so far its only) two years, but they were not attributable to the appeal of TBF, which plays only a minor part in the program. In New York, Urban Education aid remained more or less constant for seven years; with the passage of PSEN in 1974, funds will apparently be substantially increased -- but the number of participating districts is being increased at the same time, from 29 to virtually all of the state's more than 700 districts.

5. Efforts to prevent a disincentive effect and the manipulation of test scores have not been entirely auspicious. Michigan's approach, to provide a counter-incentive by requiring score increases to maintain the funding level, ran into the problem that has already been mentioned: It proved politically impossible to reduce state aid when scores did not increase. Furthermore, this counter-incentive may necessitate the segregation of low-scoring pupils (so that teachers can pinpoint score-raising actions to them) as well as the administration of an extra test to avoid a regression

effect (which in turn raises a question about whether target pupils have been properly designated). The device of giving the aid-determining test at an early grade, as suggested by the Fleischmann Commission for New York, might minimize the disincentive over the long run but was rejected by the NYDE on the ground that it would be "unfair" to the local districts, which may be interpreted to mean that it would encourage manipulation in the form of an inducement for schools to choose instructional methods that would produce the lowest scores at early grades.

California's method in ECE, requiring that a district initiate new programs in a limited number of grades or schools and then be permitted to install them elsewhere only upon demonstration that they have been successful, seems to offer the most promise, but it is not without problems, either. Like Chapter 3 in Michigan, it offers a counter-incentive, increased funds for expansion of new programs, yet without the threat -- or at least without the immediate threat -- of a reduction in state aid should the programs prove unsuccessful.[20] At the same time, California probably has

---

[20]At first glance, the procedure might seem to raise a problem analagous to the dilemma discussed above: If a district is required to "start small," there are likely to be a number of low-scoring children who will not be served by the new programs even though their low scores indicate they have extra "needs," and if the district is not granted expansion funds, they will continue to be unserved. However, a reasonable response to this objection would be that the failure of a district to receive expansion funds implies that it has not yet found reliable ways of improving performance, so the unserved children are not being deprived of beneficial programs. The implementation of ECE sought to keep down the number of unserved low-scoring children by requiring that half the participating schools in a district be among those with the largest numbers of low-scoring pupils.

reduced the temptation to manipulate scores simply by having the amount of funds depend on the scores to only a minor extent. The flaws in these approaches are that they put a heavy burden on the state in deciding whether programs have met the criteria of success (and in some states this would be regarded as intolerable infringement on local autonomy); and that, to the extent that funding does not vary with scores, the presumed benefits of TBF would not be forthcoming.

CHAPTER 3: ISSUES FOR FURTHER RESEARCH

There are a number of questions connected with TBF on which the experience to date for one reason or another casts no light at all. We shall consider some of them in this chapter, bringing to bear both logical analysis and such pertinent literature as there may be, and suggesting lines of future inquiry which may improve our understanding of the premises and the implications of TBF.

Entry testing

One way of seeking to avoid the disincentive effect that has not been tried is to administer the funding test very early in the school career, or even before it begins, so that school experience could not depress the scores. Thus, the Fleischmann Commission urged that

> "learning readiness" tests that are statistically valid and accurate be introduced as soon as possible into the New York State testing pattern, so that resource distributions to elementary students can be directly related to their capacity to learn as measured when they first enter school. Early testing is essential to see that the resource-distribution mechanism does not operate as an incentive for poor performance or in such a way that districts lose funds as they help their students to perform better (Fleischmann Report, vol. 1, p. 68).

The Commission's reason for not recommending that this be done immediately was that such tests were not yet available (and consequently, as we have seen, it proposed the interim use of a third-grade test).

Similarly, in his TBF proposal to the California State Senate, Benson (1972, p. 40) suggested the use of first-grade scores,

because "if tests at later grade levels are used as a criterion for disbursement of funds, they may act as a disincentive to teachers and administrators to improve student achievement." But he acknowledged that this might simply alter the effect of the disincentive rather than eliminate it: "Even the use of grade one test scores might act as a disincentive to schools to offer pre-school educational services," which might be highly beneficial to some children.[1] He could also have said, though he did not, that schools might be able to improve the performance of young children by strengthening their programs of parental education, and TBF based on testing at school entry might deter them from that as well. Furthermore, entry testing might introduce a disincentive effect into the home, since taxpaying parents would have to decide whether to provide a favorable early educational environment for their children at the risk of reducing outside aid for the school district and thereby possibly increasing their taxes. A decision not to provide a favorable environment

---

[1]Benson nevertheless proposed that grade one test scores be used "[i]n spite of the shortcomings." He added that "if the state wishes to provide pre-school educational services to potential low achievers, it can specify that a certain portion of compensatory funds allocated to a district be used for that purpose" -- i.e., attempting to control the disincentive by limiting the uses to which the funds could be put; and that "grade nine test scores should be the basis for the allocation of compensatory funds among high school students," on the dubious ground that "[s]ince elementary school teachers are not likely to suffer if compensatory funds received by the high school are cut, there should be very little disincentive to improve student performance at the elementary level" (Benson, 1972, p. 40).

could be justified, consciously or otherwise, by the expectation that the schools would "make up" for it, the more so if they received additional funds for doing so.

Since entry testing has not been tried as a basis for TBF, there is no empirical evidence to indicate which if any of these possible results would ensue. It should not be difficult to devise experiments which would furnish that evidence.

More research is needed, too, on the susceptibility of TBF to score manipulation. It should be pointed out that manipulation would not actually require an elaborate conspiracy, and again perhaps not even a conscious intent. Selection of a test level that was too difficult for some students, or a promotion policy that tended to place students in a grade they were not ready for, would have the effect of producing scores that would be lower than they might be otherwise; or teachers, in their role as test administrators, might give less than clear instructions, permit noise and other distractions while the test was being given, or shave a few minutes off the time allowed to students to take it. (See also Stake, 1974, pp. 68-75.) Having the test administered by the state or by "an outside agent" instead, as suggested by the Fleischmann Commission (Fleischmann Report, vol. 1, p. 68), might add considerably to the expense of the procedure, and it would almost certainly arouse resentment on the part of teachers and local school administrators, who would interpret it -- correctly -- as a sign that they were not trusted. It is

noteworthy that, in all three states whose experience we have studied, the tests for TBF were locally administered.

None of the three states has collected any evidence to show how serious the disincentive effect or score manipulation really are or might be. Perhaps the concern about them is unnecessary. There is, after all, a countervailing force: the desire of teachers and administrators to produce high scores in order to demonstrate to their communities that they are doing their jobs well. Communities might indeed be more pleased by rising test scores than they would be dismayed by declining financial assistance. On the other hand, particularly in districts with substantial numbers of low-SES minority-group families but with power in the hands of the majority group -- a common enough situation -- educators might evolve a strategy of "compartmentalization": preserving a group of low-performing students as a basis for additional funding while striving to improve the performance of the rest of the student body as a demonstration of their competence. Unhappily, the culture provides a convenient rationalization for such behavior. The low-performing students might well be those from minority families, whose failures educators could attribute to deficient family backgrounds over which they had no control. Once again, experiments with TBF should show whether, or to what degree, any of the possible consequences of TBF that have been mentioned here would actually occur in practice.

Criterion-referenced testing

Although the effects of using various types of tests is another issue demanding more evidence, there appears to be little reason to expect that criterion-referenced tests[2] would solve any of the problems in TBF that might arise from the use of norm-referenced tests. Validity and especially reliability are the more questionable in criterion-referenced tests; there is even considerable lack of clarity about how they are to be measured (Millman, 1973; Popham and Husek, 1971). The question of cultural bias in the funding test, as we have suggested before (see above, pp. 20-21), might not be raised in a TBF program; but if it were, there is nothing in the nature of criterion-referenced tests or test items that makes them less vulnerable to this fault than are norm-referenced tests or test items. (For a criticism of the possible bias in Michigan's statewide criterion-referenced test, see House and others, 1974, p. 14.)

Furthermore, criterion-referenced tests probably would not prevent the very thing whose prevention is supposed to be one of their main purposes: invidious comparisons among children. Nominally, a criterion-referenced test score compares a child's performance to an "objective standard" rather than to another child's performance; but it is easy enough to point out that one pupil has met the

---

[2]We use this generic term to include what are sometimes distinguished as domain-referenced, objectives-referenced, decision-referenced, treatment-referenced, and content-referenced tests (Sherman and Zieky, 1974, pp. 4-8).

criterion standard while his or her classmate has not. Invidious comparisons can be made among school districts and states as well, by pointing out, for example, that 75 percent of the third-graders in one have met the third-grade criterion in reading as against 50 percent in another. TBF actually promotes such comparisons, by making public the proportion of children in an area who are classified as having "special needs" or as being "educationally disadvantaged" by the test criterion (Madaus and Elmore, 1973, p. 3016).

In fact, as these observations begin to indicate, criterion-referenced tests are not as different from norm-referenced tests as they might seem to be from the definitional distinctions that are sometimes made. They cannot be told apart "by simple inspection" (Glaser and Nitko, 1971, p. 654), and it may even be misleading to classify a test as being either one or the other (Hambleton and Novick, 1973, p. 162). Perhaps the most generally acceptable distinction would be in terms of the way in which scores acquire meaning: norm-referencing gives meaning to a performance score by relating it to the distribution of scores obtained by a defined group of people, the "norming" group; criterion-referencing gives meaning to a score by relating it to what would be a perfect score on a specified performance standard, the "criterion" (cf. Glaser and Nitko, 1971, p. 653; Messick, 1975, p. 957). Thus, it would be more accurate to speak of "criterion-referenced interpretations" and "norm-referenced interpretations" than of two types of tests.[3]

[3]The "norm" to which reference is made is sometimes alluded to as if it were necessarily the "average" or mean score, but actually it may be any point in the distribution of scores. Hence, it might also be better to speak of "distribution-referenced" than of "norm-referenced" interpretations. But in recognition of the clumsiness of that phrase, and in deference to customary usage, we shall continue to use "norm-referenced."

However, any criterion must ultimately rest on some sort of judgment (Millman, 1973, p. 206); the judgment is bound to take into account what can "realistically" be expected (Miller, 1974, pp. 34-35) or what would be "useful" for a later situation (Millman, 1973, pp. 208-210), and "realism" and "usefulness" are inextricably tied to a perception of the normal performance of a reference group.

This relationship is sometimes obscured by expressing a criterion in terms of what children of a given age or grade should be able to do rather than in terms of what they actually do. But it would be futile to insist that children in a certain grade "should" be able to do what most of them cannot do; no one would propose, for example, that the criterion of mathematics performance for first-graders should be the ability to solve quadratic equations. Conversely, there is nothing to be gained by setting a criterion that the vast majority of those to be tested could easily meet -- e.g., for twelfth-graders, the ability to add two single-digit numbers. Pursuit of this line of reasoning leads to the conclusion that a criterion, to be meaningful, must bear some relationship to the actual distribution of performance scores among children who are like those to be tested.

Perhaps it is not too much to say that the chief difference between a criterion-referenced and a norm-referenced interpretation is that, in the former, scores are related to a norm that is often unarticulated, is arrived at impressionistically, and is called a "criterion," while in the latter, scores are related to a norm that

is explicit, is empirically derived, and is called a "norm." Or, in the more moderate language of Sherman and Zieky (1974, p. 19), in a test intended for criterion-referenced interpretation the decision about the norm "will often precede or occur simultaneously with the generation of [test] material, whereas it typically follows the generation of material in a norm-referenced sequence." Even this difference can disappear when, as is quite possible, a criterion is defined straightforwardly as a point in the distribution of scores among a specified group (Fremer, 1972; see also Hambleton and Novick, 1973, pp. 161-162). We saw this happen in Michigan, where the "performance objectives" of the TBF legislation were translated by the MDE into "one year's [normal] growth." It also occurred in New York when the PEP tests were first developed; the "minimal competence level" -- a criterion score -- "was set to correspond to the percentile closest to the third stanine division of the distributions, which happened to be the 23rd percentile" (Fleischmann Report, vol. 1, p. 29). Nor is there any reason in principle why normative information cannot be supplied about the scores on a test designated as "criterion-referenced," and some test publishers do so (Knapp, 1974).

There are decision-making situations in which a criterion-referenced interpretation of a test score is more appropriate than a norm-referenced interpretation, but consideration of the nature of those situations indicates that, oddly enough, the allocation of funds is just the sort of decision for which criterion-referencing

is probably not appropriate. As Hambleton and Novick (1973, pp. 162-163) point out, "Norm-referenced measurement is particularly useful in situations where one is interested in 'fixed-quota' selection," because the percentile in a score distribution can be set to correspond to the quota being sought. Criterion-referencing, on the other hand, is useful where there is "a 'quota-free' selection problem" -- that is, where there is no need to limit "the number of individuals who can exceed the cut-off score or threshold ..."[4] If test scores are to be used as the basis for allocation of funds, some sort of quota would seem to be required, in order to allow control, or at least estimation, of the size of the monetary appropriations that will be needed. In other words, it is difficult to see how TBF could be successfully implemented without declaring in advance the proportion of the pupil population to be defined as having extra educational needs, and the score at which that proportion is reached is, in effect, a norm.

Finally, the use of criterion-referenced score interpretations for TBF purposes could well prove to be self-defeating. We suggested earlier (see above, pp. 21-22) that the impetus for proposing that TBF rely on criterion-referenced "tests" apparently stemmed from the hope of avoiding the criticisms that have been directed at norm-referenced tests. Holmen and Docter (1974, p. 60), however, make the following plausible assertion:

---

[4]For a similar statement, see Glaser and Nitko (1971), p. 655.

> At the heart of the criticisms about tests and testing is one fact ... Tests are often used as tools for the allocation of limited resources or opportunities. Put another way, educational and psychological tests are frequently designed to measure differences among individuals so that one person receives a reward or privilege which another is then denied .... Tests, therefore, are likely to stir strong emotions ...

While TBF is not primarily concerned with measuring differences among individuals, it does obviously entail the allocation of limited resources among states or school districts, some of which will receive more money than others, and consequently it can reasonably be expected that the test upon which it depends -- whatever it may be called -- will "stir strong emotions" and arouse the same kind of opposition now aroused by norm-referenced tests. In that connection, it is worth observing that, no matter how criterion-referenced tests or test scores may be construed, they would have many of the same attributes and potentialities that have been the objects of criticism of norm-referenced tests -- that they are impeffect predictors, that they are often assumed to measure innate characteristics, that they may influence teachers' expectations of their pupils, that they imply there is a single right answer to every question, that they tend to restrict educational change and flexibility, that they can distort an individual's self-concept and level of aspiration, that they may be used to organize homogeneous learning groups, and that they invade privacy (Holmen and Docter, 1974).

The assumptions of TBF

Several of the assumptions associated with TBF can be questioned, and it would seem well to explore them before embarking on larger-scale

trials of the concept. One of these is that funds for compensatory education should be targeted precisely and exclusively to those pupils who have the need for them, that to do otherwise amounts to "resource leakage" and is thus wasteful. Aside from the difficulty of doing this, which has already been mentioned (see above, pp. 66-67), it can be argued that the learning of low-performing pupils may be enhanced when they are instructed together with higher-performing pupils (for a review of the evidence, see Findley and Bryan, [1971]). If that is true, the use of compensatory funds in such "integrated" classrooms cannot properly be called "resource leakage"; it would simply be part of the cost of compensatory education. From this point of view, the provision of New York's PSEN program, that up to 30 percent of the funds could be spent on otherwise ineligible pupils where necessary to avoid performance-segregated classes, would be quite justified in principle.[5]

Another questionable assumption is that, because TBF is related directly to a measure of educational outcomes, it is more likely to

---

[5]The same point could be made about the segregation of low-SES pupils in IBF programs, but it is in a sense unnecessary to do so, because that kind of segregation is generally regarded as undesirable, anyway, by virtue of its social implications. Thus, the guidelines for Title I projects in California included the provision that "Title I projects will not be approved if they (1) create special tracks for the educationally disadvantaged [who, by existing Title I rules, are low-SES children]; (2) establish adjustment, pre-grade or junior grade classes for the educationally disadvantaged; (3) isolate Title I children from the mainstream of school life for a period of time greater than one-half of the regular school day" (quoted in Berke and Kirst, 1972, p. 107). Because of the relationship between test scores and SES, segregation by ability would often result in a high degree of segregation by SES.

bring about the equalization of those outcomes than is IBF. An analogy with experience in public health would suggest instead that dealing with low test scores after they have manifested themselves may be a less reliable way of improving them than identifying the circumstances which lead to low scores and counteracting their effects before the effects appear. These circumstances are the ones that give rise to the need for additional or compensatory educational resources; and if they include the circumstances encompassed in the notion of low SES, then the second strategy describes the rationale underlying the IBF approach.

What this implies is that the assumption which is at the heart of the TBF concept — viz., that test scores are the true indicators of the need for additional education resources -- also should be carefully examined. A closer look at it does indeed evoke doubt about its validity.

In much of the recent educational literature, an educational need has been defined in terms such as "a condition in which there is a discrepancy between an acceptable state of affairs and an observed state of affairs" (Anderson and others, 1975, p. 254), or, more simply, as a "deficit" (Glass, 1970, p. 46). This definition is consistent with, and perhaps has even helped to produce, the TBF idea. If an average performance score is "an acceptable state of affairs," then a below-average score is ipso facto a discrepancy or deficit, and hence an indicator of need. Conversely, an average score would be

taken to signify the lack of a discrepancy and thus the absence of a need. A couple of hypothetical illustrations may serve to show that this definition leads to conclusions that few people would be willing to accept.

Suppose that a test of reading comprehension were administered to 1,000 ten-year-old deaf children. The great majority of them would score considerably below the average for ten-year-old children with normal hearing, since the inability to hear is a serious handicap in learning how to read. But suppose -- and it is not at all unlikely -- that a few of them, say five, performed at the average. One would hardly conclude that deafness had not been a handicap for these five children and that they therefore had no need for extra educational resources. Either of two other inferences would be more plausible. (1) These children had already been getting extra educational resources in a highly effective way, and that is what enabled them to achieve an average score. (2) If they had not been getting extra resources, they had extraordinary capabilities which had enabled them to overcome the effects of their handicap so far -- and if they were now to receive that additional help, they could presumably do even better. In either case, to deprive these five children of extra help because of their average scores would surely be widely regarded as an injustice to them.

For a second illustration, suppose that a test of achievement in mathematics were given to two groups of eighth-grade children.

One group lived three miles or more from the nearest school, the other lived within a block of the school they attended; the groups were otherwise similar. Suppose -- and again, it is not an unreasonable supposition -- that the mean score for the two groups were the same. Here too, it would be unwarranted to conclude that the children who lived far from school had no need for transportation services. Rather, it was probably the fact that they had been transported to school which made their average achievement in mathematics possible; or, if they had not been transported, they had unusual abilities which enabled them to learn mathematics on their own. Perhaps they had the help of skilled parents, but few people would contend that children who live far from school ought to be required to rely exclusively on their parents for their education. If schools help children at all (or if schooling is compulsory), the principle of equality demands that children who live far from one should be brought there -- even if, after they get there, they perform as well as their classmates who walked.

Living at a great distance from school is a handicap; we might call it the handicap of distance (or "impaired access"). It is different from deafness in that it is a socially imposed and a socially removable handicap. Society has the option, so to speak, of building a school within walking distance of every child, but it has decided that that is too expensive in areas of sparse population. Consequently, it has "under-invested" in the education of children living in such

areas, and the principle of equality places society under an obligation to make up for that under-investment. It does so by transporting these children to school at public expense.

The definition of "need" which is implicit in these illustrations is "a requirement for extra resources to overcome or mitigate the effects of a condition adverse to learning." Deafness and distance are such conditions; to say that a child suffers from a handicap is to say that the child is in need of special help. If a handicapped child displays no performance deficit, it does not follow that the child has no unusual need; the inference, rather, is that the need, the requirement for extra resources, has at least partially been met.

Coming from a family of low SES may be viewed as another type of handicap. To repeat what was said at the outset of this study (see above, p. 3), low-SES children are likely to have suffered from such educationally relevant disadvantages as "inadequate nutrition, clothing, health care, and shelter and space; a restricted range of experiences; and peers and parents who provide little help in verbal development or other scholastic skills and who do not offer models of high aspirations." In short, these children, compared to others of higher SES, have had relatively little invested in them. Adapting a term employed by Levin and others (1971, pp. 198-199) and by Benson (1972, p. 39), we may call this the handicap of "deficient capital embodiment."

When seen in this light, low SES would not be a "proxy" measure of educational need; it would be the direct source of a need. The generally low test scores of low-SES children would signify the existence of a need, but they would be interpreted as the symptom of a widespread condition, not as the condition itself. If some low-SES children achieve average (or "non-low") scores, the conclusion that they have no special needs would be overly hasty. One of the more likely meanings -- just as in the case of the children transported to school -- is that their needs have so far been met, by the application of extra resources to compensate for their deficient capital embodiment. Justice and equality would seem to require that these extra resources continue to be supplied.[6] To say that "there would be no need for compensatory education if the correlation between socioeconomic status and achievement were reduced to zero" (Garms and Kelly, 1970, p. B11) would be analogous to saying that there would be no need for transportation services if the correlation between distance from school and achievement were reduced to zero; the defect in the reasoning is that it may be the transportation services which are responsible for the reduction of the correlation.

The same sort of logical slip may account for the interpretation made of the previously cited finding by Glass (1970; see above, pp. 8-9)

---

[6]We assume here that additional resources can, if applied properly, improve learning. This is by no means an uncontested assumption, but the weight of the evidence favors it; for some pertinent summaries, see Benson (1972), pp. 29-33; Garms and Kelly (1970), pp. B7-B10; Guthrie and others (1971), pp. 57-91. If additional resources cannot improve learning, it would be difficult to make a case for either IBF or TBF.

that about one-third of the children in Title I schools had "no critical needs for compensatory programs." Imagine a teacher in charge of a compensatory reading program for 30 pupils. If ten of these pupils are performing at an average level, and the teacher feels they will continue to do so, the teacher might well say that they have "no critical needs" for further participation. But that could simply reflect the teacher's belief that the program they are in, and the programs which preceded it, have met their needs (i.e., have been sufficient to compensate for deficient capital embodiment). A reduction in funds for that school would be called for only if there were fewer pupils like these coming after them -- that is, only if there were a change in the socioeconomic composition of its student body.

If some low-SES children have average scores and yet have not had extra resources devoted to their education, it still does not necessarily mean that they have no need for those resources. One other possible interpretation is that the schools they have attended have concentrated whatever resources they do have on instruction in reading and mathematics, the subject areas usually being referred to when "scores" are under consideration. If so, extra resources would be required by their schools to improve their education in the areas that have thus been neglected -- science, social studies, foreign languages, and so on. Full equality is surely not restricted to equality in reading and mathematics; low-SES children ought to

have the same opportunity as high-SES children to raise their performance in other subjects as well.

Another possibility, as in the case of other handicapped children, is that low-SES children with average scores have unusual abilities which enabled them to overcome their handicap without additional help. (This would also be the interpretation if the test had been one given upon or before entry into school.) Presumably, if such children were provided with additional resources, their scores would be even higher, and that result probably should be regarded as a necessary element of equality. Educational equality can hardly be said to be complete if a high-SES child of high ability reaches a high level of achievement while a low-SES child of the same high ability reaches only an average level of achievement. Full equality would require that children of the same ability be helped to reach the same level of achievement, without regard to SES.[7]

It begins to appear, therefore, that the absence of a low score may not be a sign of the absence of a need for compensatory educational resources; and if that is so, the basic assumption of TBF, that test scores are the true need indicators, is called into question. Interestingly enough, even some of the proposals for TBF seem to contain doubts about it. We noted earlier (see above, p. 18) the fear expressed by the Fleischmann Commission that, if allocations

---

[7] This is not meant, of course, to deprecate the importance of raising below-average scores to average scores, but only to say that that is insufficient; it does not amount to the realization of full educational equality.

were determined by scores on a test given subsequent to school entry, districts would lose funds "just at the time when they ... have discovered how to use them effectively." Implicit in that concern is the presumption that rising test scores would mean not that needs were disappearing but that they were continuing and were being met. In H. R. 5163, allocations would have been based on the number of low-scoring children and on the number of children who were students in state-supported schools for handicapped or for neglected or delinquent children or who were children of migratory agricultural workers. Test scores of these latter categories of children were not to be taken into account, which implies that their needs could not be measured by their scores. It would not seem to be far-fetched to think that the same reasoning could be applied to other children whose handicaps may be less overt.

One reason why these doubts have not become more prominent may be that, as was suggested above, they have been suppressed by the "deficit" or "discrepancy" conception of educational need. Another, closely related reason, which perhaps particularly explains why they have not surfaced during the implementation of the state TBF programs, may be that children who are performing at the average seem not to be "problems." They are "doing well enough by ordinary standards" (Benson and others, 1974, p. 85), and no one knows whether they could do better. To an educational policy-maker or administrator struggling with a tight budget, it can easily seem a correct course

of action not to continue to devote additional resources to the education of average-performing children when there are so many children who are performing below average.

This route to TBF, however, rests on at least three assumptions: (1) that average performance is "good enough" for low-SES children; (2) that a reduction of resources for these children will not lead to a decline in their performance; and (3) that high-SES children with low test scores need additional educational resources. The first two of these have already been touched on, but the third bears further consideration.

The beneficiaries of a change from IBF to TBF would be high-SES children with low test scores (or, more precisely, the states or districts in which such children live). It should be clear that, under either IBF or TBF, compensatory funds would continue to flow to low-SES children with low scores, since they would meet the criteria under either procedure. But under IBF, funds are also allocated on behalf of low-SES children who do not have low scores; these children would not meet the criteria of TBF, whereas high-SES children with low scores would, and so funds would be diverted from the former to the latter.[8]

---

[8]The net effect on a state's or district's funding would depend on the relative sizes of the appropriations for IBF and TBF, the proportion of children classified as "low performers," the relationship between that proportion in the state or district and the proportion estimated as being of low SES under the IBF provisions of Title I or the state compensatory-education program, and, in the case of state-level TBF, the other provisions of the state aid formula (Levin and others, 1972, pp. 205-206; Berke, 1974, p. 125). For an initial effort at calculating the effects, among and within a few states, of basing Title I allocations on test scores, with definitions of "low performers" at various points in the score distributions, see Guthrie and others (1974). However, it is not the magnitude of the effect which is of concern here, but only its direction, which of necessity is away from low-SES areas and toward high-SES areas.

The question then arises as to the likelihood that the low scores of high-SES children will be raised by the infusion of additional funds into the districts where they live.

The answer is far from certain. In the first place, high-SES families vary in their value preferences. Those with low-scoring children may have chosen to invest less than other families in their children's preparation for and success in school, and it cannot be taken for granted that supplying additional funds to the districts in which they live would alter their preferences or their behavior.

But suppose that the schools could make up for this under-investment, as we have assumed they could for under-investment among low-income families.[9] Some analyses (e.g., Wade and others, 1975) show that the median family income of a school district is virtually unrelated to its per-pupil value of taxable property, which is the major source of local revenue for educational expenditures. Insofar as this is so, additional funds allocated on behalf of children with low test scores would permit high-SES districts to increase their expenditures, with the presumed effect of raising scores. Other analyses, however (e.g., Berke, 1972), show that high-income families tend to live in districts that already have relatively high per-pupil expenditures, and in that case, it may be doubted that enabling

---

[9] See footnote 6 above.

these districts to increase their expenditures still more would bring about higher scores among their pupils.[10] It is conceivable that, in such districts, low-scoring children have had the wrong resources applied to their education; the condition adverse to their learning has been a type of teaching style or curriculum or instructional material or other component of schooling that has not been suitable to their particular characteristics, and additional funds would permit their schools to make a thorough search for more suitable resources. But if that is so, then all districts need more funds, not just those with a relatively high-SES population composition, since the same kind of error is presumably made just as often in the case of low-SES children.

All that can be said with confidence, then, is that before assertions can be made about the potential effect of increased funding on the performance of low-scoring children of high-SES background, more must be learned about the reasons for the low scores, and about the extent to which TBF would stimulate increased expenditures for schooling in high-SES districts.

The data base for IBF

The ambiguities that have become evident in TBF prompt a reconsideration of IBF, to see if its deficiencies can be overcome

---

[10] It is also questionable whether the allocation of additional funds to high-expenditure districts, particularly if this is done at the expense of funds for low-expenditure districts, would meet judicial standards of equity.

short of abandoning it altogether. While a comprehensive reconsideration is beyond the scope of this study a few observations may be in order.

The problem of keeping the data base for IBF up-to-date may not be as serious as has been suggested. Determining fund allocations "as if a 21-year-old man or woman were still in the second grade" (Quie, 1973, p. 3) would not be so objectionable if the 21-year-old person for whom the funds were being provided in a given year had since been replaced by a second-grader of the same SES. Major changes in the socioeconomic composition of school districts probably do not take place in periods of less than a few years. Major changes in the relative proportions of low-SES people among districts probably take even longer; if so, a district with greater needs for educational resources would continue to receive larger grants than a district with lesser needs long after the original data were collected. This is a conjecture that could and should be empirically investigated. If current data are important, it might be just as easy for the funding agency to conduct a biennial sample survey of socioeconomic status in each district as it would be for the agency to administer a test biennially to a sample of pupils, and it would almost certainly be less expensive.

Although we have been using the term "income-based funding" for procedures in which allocations vary with SES, a question may be raised about whether income -- which is a proxy for SES — is the

most satisfactory measure of the many facets of that condition. Beside the issue of invasion of privacy, the educationally relevant aspects of the family environment are not necessarily tapped by the size of the family income. A recent comprehensive review of research on the determinants of status (Haller and Portes, 1973) led its authors to conclude that "the status fate of youth may well hinge more on the psychological than on the economic support [parents] are able to give them."

Perhaps a better measure of SES for purposes of allocating compensatory-education funds would be years of parental schooling. Data gathered from 45 elementary schools in New York state (Garms and Smith, 1969) showed that average number of years of father's schooling, or of mother's where father was absent, could account for nearly 45 percent of the variance in third-grade test scores (sum of the percent scoring below the fourth stanine in reading and the percent scoring below the fourth stanine in arithmetic).[11] In a later application of the same methodology to a larger and more representative sample of schools in the same state (Garms and Goettel, 1972), virtually the same proportion of test-score variance could be

---

[11]Citation of this kind of evidence in support of a measure of need does not constitute a return to the assumptions of TBF. As has been pointed out, low test scores can properly be regarded as a symptom of educational need; the problems with TBF are that it seems to take the symptom for the source and to misinterpret the absence of low scores. What is being suggested here is that the source of educational needs can legitimately be sought among the correlates of low test scores.

accounted for by either average years of father's schooling or average years of mother's schooling.[12] This proportion is rather less than the 65 percent of variance that could be accounted for by all 30 variables that were used, and less even than the 62 percent that could be accounted for by combinations of only the three most efficient variables (percent of pupils from broken homes, percent living in overcrowded housing, and average years of mother's education); but what parental education alone lacks in predictive power might be made up for, in terms of its use as a fund-allocation instrument, by its simplicity and precision of definition, objectivity, availability, and relative innocuousness as a subject of inquiry when necessary. How much difference in the distribution of funds would be caused by various measures of SES is, again, a matter subject to empirical investigation.

If parental education should prove to be an acceptable measure of SES, a method is available for converting it into a new form of IBF which would be consistent with the concept of low SES as a handicap that was presented above. Dugan (1969) has estimated the amounts of capital embodiment in children that are represented by different levels of parental education. The estimated amounts, based on a measure of foregone income, generally accounted for 80 to 95 percent of the variance in pupils' verbal-skill scores,

---

[12]The relevant data are in Appendix D of Berke, Campbell, and Goettel (1972). It is interesting to note that in the later study, the assertion made in the earlier study, that SES was but a "second alternative" to test scores, was not repeated.

the score data being drawn from the Equal Educational Opportunities Survey (Coleman and others, 1966). Furthermore, by calculation of marginal rates of substitution between "parental educational services" and "school resources" (per-pupil expenditures, by state), he arrived at "an indication of the amount of school resources required to elevate a culturally deprived child to an achievement level comparable to the national norm" (Dugan, 1969, p. 146). With refinements, this technique might be used to develop ratio indices for a funding formula (for illustrations, see Levin and others, 1971).

It might be objected that, if funding were determined by parental educational levels, school districts might be discouraged from trying to raise those levels, as they might otherwise do by, for example, offering high-school equivalency courses, establishing a community college, or providing other kinds of continuing education for adults -- i.e., another form of the disincentive effect. But if increased parental education enhances children's aptitude and appetite for learning, most educators would probably welcome it as a factor making their jobs easier and more satisfying. The decline in funds that would result ought not cause dismay, because -- unlike the case with rising test scores -- the need for the funds would be genuinely declining simultaneously. Another concern might be the possibility of deception, in the form of under-reporting by parents of their education, but that seems much less likely than under-reporting of income. Once more, the accuracy of these speculations could be studied in experimental situations.

Summary

Analysis of some of the as yet untried solutions to problems of TBF, and of some of the assumptions associated with it, reveals areas that demand further thought and empirical investigation. Testing upon entry to school rather than in later grades may only change the way in which the disincentive effect is manifested instead of eliminating it. Manipulation of scores may be easier than it has seemed. The desire of teachers and administrators to demonstrate their competence by producing high scores might be sufficient to counteract the disincentive effect and the temptation to manipulate scores, but evidence is lacking.

There is little reason to believe that criterion-referenced tests would overcome the opposition to TBF that it is thought may arise from the use of norm-referenced tests. Criterion-referenced tests are just as subject to cultural bias and to use in invidious comparisons as are norm-referenced tests. In fact, the two are not different kinds of tests so much as different ways of interpreting test scores; and even the interpretations are not radically distinct, since a criterion is of necessity related to a distribution of scores. Moreover, criterion-referenced interpretations are especially inappropriate to decision-making situations in which the number of pupils to be placed in a category must be limited, which seems to be the case for TBF. Finally, the use of criterion-referencing for TBF purposes might well bring upon it all the criticisms that have hitherto been directed at norm-referencing.

Among the assumptions of TBF that require further exploration are that compensatory-education funds should be spent exclusively on students who have extra needs; that TBF is more likely to equalize the scores of different SES groups than IBF is; and, most importantly, that test scores are true indicators of the need for additional educational resources. The definition of "educational need" that is consistent with TBF describes it as a "discrepancy" or "deficit" relative to a desired condition such as normal performance, leading to the implication that the absence of a low score signifies the absence of need. This implication would probably be rejected by most people if it were applied to children with handicaps like deafness or residence at a great distance from school. Additional resources may have reduced the effects of their handicap, enabling them to achieve higher scores than they might otherwise; but if so, it would seem that these resources, rather than being withdrawn if the children reach the average, should continue to be provided as long as the handicap exists, in order to sustain their higher performance. Thus, a more generally acceptable definition of "need" would be "a requirement for extra resources to overcome or mitigate the effects of a condition adverse to learning."

Low SES can properly be viewed as the handicap of "deficient capital embodiment." In that light, it would itself be the source of needs, not merely a proxy measure of them. The low scores of low-SES children would be interpreted as the symptom of a widespread

need, rather than as the condition creating the need. Low-SES children with average scores would still need additional funds, to allow improvements in areas of their education other than those measured by the test scores or to enable those of high ability among them to reach high, rather than simply average, levels of achievement. These results are necessary elements of the concept of educational equality.

TBF would divert funds away from districts which have low-SES children with non-low scores toward districts which have high-SES children with low scores. Present evidence does not yield a clear answer to the question of whether these funds would have the effect of raising the scores of the children on whose behalf they would be allocated.

The doubts about TBF lead to a reconsideration of IBF. The data base for it may not become obsolete as quickly as has been thought, and it could be kept up-to-date fairly easily. Parental education may be a more satisfactory measure of SES than family income, and a technique has been developed which would relate it to amounts of capital embodiment in children. It might contain its own kind of disincentive effect and an inducement for deception, but these are probably less serious than either with an income measure or with TBF. However, these, too, are matters for further study.

REFERENCES

Anderson, Scarvia B.; Ball, Samuel; Murphy, Richard T., and associates, 1975. Encyclopedia of Educational Evaluation. San Francisco: Jossey-Bass.

Bailey, Stephen K., and Mosher, Edith K., 1968. ESEA: The Office of Education Administers a Law. Syracuse, NY: Syracuse University Press.

Benson, Charles S., and others, 1972. Final Report to The Senate Select Committee on School District Finance, vol. I. [Berkeley, CA:] (Report prepared for the California State Senate). (ED 083 718)

Benson, Charles S.; Goldfinger, Paul M.; Hoachlander, E. Gareth; and Pers, Jessica S., 1974. Planning for Educational Reform: Financial and Social Alternatives. NY: Dodd, Mead.

Berke, Joel S., 1972. "School Finance and Inequality of Educational Opportunity." Pp. 1-16 in Joel S. Berke, Alan K. Campbell, and Robert J. Goettel, Financing Equal Educational Opportunity: Alternatives for State Finance. Berkeley, CA: McCutchan.

Berke, Joel S., 1974. Answers to Inequity: An Analysis of the New School Finance. Berkeley, CA: McCutchan.

Berke, Joel S.; Campbell, Alan K.; and Goettel, Robert J., 1972. Financing Equal Educational Opportunity: Alternatives for State Finance. Berkeley, CA: McCutchan.

Berke, Joel S., and Kirst, Michael W., 1972. Federal Aid to Education: Who Benefits? Who Governs? Lexington, MA: D.C. Heath.

Burke, Arvid J.; Kelly, James A.; and Garms, Walter I., 1971. "Educational Programs for the Culturally Deprived--Need and Cost Differentials." Pp. 79-99 in Roe L. Johns, Kern Alexander, and K. Forbis Jordan, eds., Planning to Finance Education. Gainesville, FL: National Educational Finance Project.

California State Department of Education, 1973a. "The Early Childhood Education Proposal: A Master Plan to Redesign Primary Education in California." Sacramento: The Department.

California State Department of Education, 1973b. "Policies for Early Childhood Education." Sacramento: The Department.

California State Department of Education, 1974. "Early Childhood Education: First Annual Evaluation Report 1973-74." Sacramento: The Department.

[California State Department of Education], 1975. "State Department of Education Response to the Legislative Analyst's 'Concerns about the Evaluation of the Early Childhood Education Evaluation Program (3/5/75).'" [Sacramento: The Department.]

Cohen, Michael A.; Levin, Betsy; and Paver, Richard, 1973. The Political Limits to School Finance Reform. Washington: The Urban Institute. (UI 727-2)

Coleman, James S., 1971. "New Incentives for Schools." Pp. 70-90 in James W. Guthrie and Edward Wynne, eds., New Models for American Education. Englewood Cliffs, NJ: Prentice-Hall.

Coleman, James S., and others, 1966. Equality of Educational Opportunity. Washington: U.S. Government Printing Office.

Dugan, Dennis J., 1969. "The Impact of Parental and Educational Investments upon Student Achievement." Pp. 138-148 in American Statistical Association: Proceedings of the Social Statistics Section 1969. Washington: The Association.

Feldmesser, Robert A., and Echternacht, Gary J., 1975. Performance Contracting as a Strategy in Education. Princeton, NJ: Educational Testing Service. (PR-75-10)

Findley, Warren G., and Bryan, Miriam M., [1971]. Ability Grouping: 1970: Status, Impact, and Alternatives. Athens, GA: Center for Educational Improvement, University of Georgia. (ED 060 595)

The Fleischmann Report on the Quality, Cost, and Financing of Elementary and Secondary Education in New York State., 1973. 3 vols. New York: Viking.

Fremer, John J., Jr., "Criterion-Referenced Interpretations of Survey Achievement Tests." 1972. Princeton, NJ: Educational Testing Service. (TDM-72-1)

Garms, Walter I., and Goettel, Robert J., 1972. "Measuring Educational Need: Developing a Model for Predicting Achievement Levels from a Composite of Socioeconomic Characteristics." Pp. 169-191 in Joel S. Berke, Alan K. Campbell, and Robert J. Goettel, Financing Equal Educational Opportunity: Alternatives for State Finance. Berkeley, CA: McCutchan.

Garms, Walter I., and Kelly, James A., 1970. "Estimating Need for Compensatory Education through Socio-Economic Predictions of Achievement." Pp. B-1-B98 in Arvid J. Burke, James A. Kelly, and Walter I. Garms, eds., Educational Programs for the Culturally Deprived--Need and Cost Differentials. Albany, NY: School of Education, State University of New York at Albany. (National Educational Finance Project Special Study No. 3)

Garms, Walter I., and Smith, Mark C., 1969. Development of a Measure of Educational Need and Its Use in a State School Support Formula. Albany, NY: New York State Educational Conference Board. (ED 038 739)

Garms, Walter I., and Smith, Mark C., 1970. "Educational Need and its Application to State School Finance." Journal of Human Resources, V (Summer), 304-317.

Glaser, Robert, and Nitko, Anthony J., 1971. "Measurement in Learning and Instruction." Pp. 625-670 in Robert L. Thorndike, ed., Educational Measurement, 2nd ed. Washington: American Council on Education.

Glass, Gene V, and others, 1970. Data Analysis of the 1968-69 Survey of Compensatory Education. Boulder, CO: Laboratory of Educational Research, University of Colorado. (ED 057 146)

Guthrie, James W.; Frentz, Anne S.; and Mize, Rita M., 1974. The Use of Performance Criteria to Allocate Compensatory Education Funds--Fiscal Issues: Resource Redistribution Consequences of the Use of Performance Criteria. Menlo Park, CA: Stanford Research Institute. (SRI Project 2158, Part II)

Haller, Archibald O., and Portes, Alejandro, 1973. "Status Attainment Processes." Sociology of Education, 46 (Winter), 51-59.

Hambleton, Ronald K., and Novick, Melvin R., 1973. "Toward an Integration of Theory and Method for Criterion-Referenced tests." Journal of Educational Measurement, 10 (Fall), 159-170.

Holmen, Milton G., and Docter, Richard F., 1974. "Criticisms of Standardized Testing." Today's Education, 63 (January-February), 50-60.

Horst, Donald P.; Tallmadge, G. Kasten; and Wood, Christine T., 1975. A Practical Guide to Measuring Project Impact on Student Achievement. Washington: U.S. Government Printing Office.

House, Ernest; Rivers, Wendell; and Stufflebeam, Daniel, 1974. "An Assessment of the Michigan Accountability System." (ED 091 821)

Kelly, James A., 1970. "Resource Allocation and Educational Need: New York City's Public Schools." Education and Urban Society, II (May), 251-276

Knapp, Joan, 1974. "A Collection of Criterion-Referenced Tests." TM Report 31. Princeton, NJ: ERIC Clearinghouse on Tests, Measurement, & Evaluation, Educational Testing Service.

Levin, Henry M.; Guthrie, James W.; Kleindorfer, George B.; and Stout, Robert T., 1971. "Capital Embodiment: A New Approach to Paying for Schools." Pp. 196-213 in James W. Guthrie and Edward Wynne, eds., New Models for American Education. Englewood Cliffs, NJ: Prentice-Hall.

Levin, Betsy; Muller, Thomas; Scanlon, William J.; and Cohen, Michael A., 1972. Public School Finance: Present Disparities and Fiscal Alternatives. [Washington:] The Urban Institute. (URI-3500)

Lipe, Dewey, and Jung, Steven M., 1971. "Manipulating Incentives to Enhance School Learning." Review of Educational Research, 41 (October), 249-280.

Madaus, George, and Elmore, Richard F., 1973. "Allocation of Federal Compensatory Education Funds on the Basis of Pupil Achievement Test Performance." Washington, DC: Statement before the House of Representatives, General Subcommittee on Education, of the Committee on Education and Labor, June 26.

Messick, Samuel, 1975. "The Standard Problem: Meaning and Values in Measurement and Evaluation." American Psychologist, 30 (October), 955-966.

Michigan Department of Education, 1972. A Description and Evaluation of Section 3 Programs in Michigan 1971-72, Report Number 1. [Lansing:] The Department.

Michigan Department of Education, 1973. A Description and Evaluation of Section 3 Programs in Michigan 1971-72, Report Number 2. [Lansing:] Research, Evaluation and Assessment Services, Michigan Department of Education.

Michigan Department of Education, 1974. A Description and Evaluation of Chapter 3 State Compensatory Education Programs in Michigan 1972-73, Part One. [Lansing:] Research, Evaluation and Assessment Services, Michigan Department of Education.

Michigan Department of Education, 1974. A Description and Evaluation of Chapter 3 State Compensatory Education Programs in Michigan 1972-73 [Part 2]. [Lansing:] Research, Evaluation and Assessment Services, Michigan Department of Education.

Miller, Donald M., 1974. "Content, Items, Decisions: Orienting Curriculum-Assessment Surveys to Curriculum Management and Modification." Educational Technology, 14 (June), 29-34.

Millman, Jason, 1973. "Passing Scores and Test Lengths for Domain-Referenced Measures." Review of Educational Research, 43 (Spring), 205-216.

Murphy, Jerome T., and Cohen, David K., 1974. "Accountability in Education--the Michigan Experience." The Public Interest, no. 36 (Summer), 53-81.

National Advisory Council on the Education of Disadvantaged Children, 1974. Annual Report to the President and the Congress/1974--Title I: Expanding Educational Growth. Washington: The Council.

National Education Association, 1973. 1973-74 NEA Handbook for Local, State, and National Associations. Washington: The Association.

Popham, W. James, and Husek, T. R., 1971. "Implications of Criterion-Referenced Measurement." Pp. 17-37 in W. James Popham, ed., Criterion-Referenced Measurement: An Introduction. Englewood Cliffs, NJ: Educational Technology Publications.

Quie, Albert H., 1973. "Remarks and Bill Description, H. R. 5163, (To Amend Title I, ESEA)." March 5 (processed).

Quie, Albert H., 1974. "Defining Educational Policy." Pp. 1-12 in Speaking Out: Law, Education and Politics--Proceedings of the Invitational Conference on Educational Assessment and Educational Policy. Princeton, NJ: Center for Statewide Educational Assessment, Educational Testing Service.

Reischauer, Robert D., and Hartman, Robert W., 1973. Reforming School Finance. Washington: The Brookings Institution.

Sherman, Matthew, and Zieky, Michael, eds., 1974. Handbook for Conducting Task Analyses and Developing Criterion-Referenced Tests of Language Skills. Princeton, NJ: Educational Testing Service. (Project Report 74-12)

Stake, Robert E., 1974. "The Suitability of Assessment Data for Resource Allocation and Educational Evaluation." Pp. 57-82 in Speaking Out: Law, Education and Politics--Proceedings of the Invitational Conference on Educational Assessment and Educational Policy. Princeton, NJ: Center for Statewide Educational Assessment, Educational Testing Service.

Today's Education, 1974. "NEA Task Force on Testing," 63 (January-February), 54.

The University of the State of New York, 1974. Evaluation Elementary and Secondary Education: A Report to the Governor and Members of the Legislature Prepared Pursuant to Section 3602 of the Education Law. Albany: The State Education Department.

Van Fleet, Donald S., and Boardman, Gerald, 1971. "The Relationship Between Revenue Allocations and Educational Need as Reflected by Achievement Test Scores." Pp. 293-317 in Roe L. Johns, Kern Alexander, and Dewey H. Stoller, eds., Status and Impact of Educational Finance Programs. Gainesville, FL.: National Educational Finance Project.

Wade, Jerome; Sacks, Seymour; and Goettel, Robert J., 1975. "Discussion Paper: Poor Children and Property Poor School Districts." [Syracuse, NY:] Educational Finance and Governance Center, Syracuse University Research Corporation.

Wargo, Michael J.; Tallmadge, G. Kasten; Michaels, Debbra D.; Lipe, Dewey; and Morris, Sarah J., 1972. ESEA Title I: A Reanalysis and Synthesis of Evaluation Data From Fiscal Year 1965 Through 1970. Palo Alto, CA: American Institutes for Research. (ED 059 415)

Wilensky, Gail, [1972.] "Reforming the Title I distribution formula." Washington, D.C.: Office of Education (DHEW), Office of Special Concerns. (ED 074 603)

Wrightstone, J. Wayne; Hogan, Thomas P.; and Abott, Muriel M., no date. "Accountability in Education and Associated Measurement Problems." NY: Harcourt Brace Jovanovich. (Test Service Notebook 33)